NCE LIBRARY

Vol. XVII, No. 2 PSYCHOLOGICAL REVIEW PUBLICATIONS APRIL, 1934

APR - 2 1934

Journal of Experimental Psychology

18

EDITED BY
SAMUEL W. FERNBERGER, UNIVERSITY OF PENNSYLVANIA

WALTER S. HUNTER, CLARK UNIVERSITY (*Index*)
HERBERT S. LANGFELD, PRINCETON UNIVERSITY (*Review*)
EDWARD S. ROBINSON, YALE UNIVERSITY (*Bulletin*)
JOSEPH PETERSON, GEORGE PEABODY COLLEGE (*Monographs*)

CONTENTS

The Temporal Indifference Interval Determined by the Method of Mean Error: H. WOODROW 167

Singer's Definition and the Generalized Law of the Geometric Mean in Numerical Estimation: C. C. LIENAU 189

Visual Perception of Horizontal Movement: R. O. VAN WATERS 223

Over-Compensation by the Non-Preferred Hand in an Action-Current Study of Simultaneous Movements of the Fingers: M. METFESSEL AND N. D. WARREN 246

The Effect of Fatigue upon Variability of Output in Muscular Work: C. W. MANZER 257

Thresholds of Vertical Movement of the Body: H. GURNEE 270

Judgment Times of Different Psychophysical Categories: S. W. FERNBERGER, E. GLASS, I. HOFFMAN AND M. WILLIG 286

The Reliability of Nonsense-Syllable Scores: J. B. STROUD, A. F. LEHMAN AND C. MCCUE 294

A New Test of Laterality: C. VAN RIPER 305

PUBLISHED BI-MONTHLY
FOR THE AMERICAN PSYCHOLOGICAL ASSOCIATION
BY THE PSYCHOLOGICAL REVIEW COMPANY
PRINCE AND LEMON STS., LANCASTER, PA.
AND PRINCETON, N. J.

Entered as second-class matter, July 1, 1920, at the post-office at Lancaster, Pa.

PUBLICATIONS

OF THE

AMERICAN PSYCHOLOGICAL ASSOCIATION

EDITED BY

S. W. FERNBERGER, UNIVERSITY OF PENNSYLVANIA (*J. Exper. Psych.*)
WALTER S. HUNTER, CLARK UNIVERSITY (*Index and Abstracts*)
HENRY T. MOORE, SKIDMORE COLLEGE (*J. Abn. and Soc. Psychol.*)
HERBERT S. LANGFELD, PRINCETON UNIVERSITY (*Review*)
EDWARD S. ROBINSON, YALE UNIVERSITY (*Bulletin*)
JOSEPH PETERSON, GEORGE PEABODY COLLEGE (*Monographs*)

HERBERT S. LANGFELD, Business Editor

PSYCHOLOGICAL REVIEW

containing original contributions only, appears bi-monthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 540 pages.

PSYCHOLOGICAL BULLETIN

containing critical reviews of books and articles, psychological news and notes, university notices, and announcements, appears monthly (10 numbers), the annual volume comprising about 720 pages. Special issues of the BULLETIN consist of general reviews of recent work in some department of psychology.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

containing original contributions of an experimental character, appears bi-monthly, February, April, June, August, October, and December, the six numbers comprising a volume of about 900 pages.

PSYCHOLOGICAL INDEX

is a compendious bibliography of books, monographs, and articles upon psychological and cognate topics that have appeared during the year. The INDEX is issued annually in June, and may be subscribed for in connection with the periodicals above, or purchased separately.

PSYCHOLOGICAL ABSTRACTS

appears monthly, the twelve numbers and an index supplement making a volume of about 700 pages. The journal is devoted to the publication of non-critical abstracts of the world's literature in psychology and closely related subjects.

PSYCHOLOGICAL MONOGRAPHS

consists of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The MONOGRAPHS appear at irregular intervals and are gathered into volumes of about 500 pages.

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

appears quarterly, April, July, October, January, the four numbers comprising a volume of 448 pages. The journal contains original contributions in the field of abnormal and social psychology, reviews, notes and news.

ANNUAL SUBSCRIPTION RATES

Review: \$5.50 (Foreign, \$5.75). **Index:** \$4.00 per volume.
Journal: \$7.00 (Foreign, \$7.25). **Monographs:** \$6.00 per volume (Foreign, \$6.30).
Bulletin: \$6.00 (Foreign, \$6.25). **Abstracts:** \$6.00 (Foreign, \$6.25).
Abnormal and Social: \$5.00 (Foreign, \$5.25). Single copies \$1.50.
Current numbers: **Journal,** \$1.25; **Review,** \$1.00; **Abstracts,** 75c; **Bulletin,** 60c.

COMBINATION RATES

Review and Bulletin: \$10.00 (Foreign, \$10.50).
Review and J. Exp.: \$11.00 (Foreign, \$11.50).
Bulletin and J. Exp.: \$12.00 (Foreign, \$12.50).
Review, Bulletin, and J. Exp.: \$16.00 (Foreign, \$16.75).
Review, Bulletin, J. Exp., and Index: \$19.00 (Foreign, \$19.75).

Subscriptions, orders, and business communications should be sent to the

PSYCHOLOGICAL REVIEW COMPANY

PRINCETON, NEW JERSEY

Journal of
Experimental Psychology

VOL. XVII, No. 2 APRIL, 1934

THE TEMPORAL INDIFFERENCE INTERVAL DETERMINED BY THE METHOD OF MEAN ERROR

BY HERBERT WOODROW

University of Illinois

The data to be presented, it is believed, at last establish within quite narrow limits and with adequate reliability the length of the indifference interval in the case of reproductions of empty intervals bounded by instantaneous sounds. This outcome appears to be due to the fact that this investigation differed from all preceding ones in that it made use of a relatively large number of subjects (two hundred and eighty) and in that the work with any one subject was confined to reproductions of just one length of interval, thus eliminating the influence of the sittings with one interval upon those with another interval.

It is an astonishing fact that seldom, if ever, have any two of the many investigators who have sought to determine the length of the temporal indifference interval reached the same conclusion. In the case of the comparison of empty intervals bounded by instantaneous sounds, the indifference intervals reported vary from 5.0 sec to less than .5 sec. Mehner,[1] using intervals from .7 to 12.0 sec, found the errors to be negative, with a few exceptions, up to 5.0 sec, but for intervals beyond

[1] *Philos. Stud.*, 2, 1885, pp. 555–556, Table II.

5.0 sec he found only positive[2] errors. His data, then, show an indifference point at 5.0 sec, but the indifference point is one at which, as the intervals are increased, the errors change from negative to positive. Estel[3] found no such change. Using intervals from 1.8 sec to 8.0 sec, he obtained only negative errors throughout the entire range. Far from changing to positive at 5.0 sec, beyond this point the negative errors simply became *larger* negative errors. Hülser,[4] using only one interval in this range, namely 2.0 sec, found the error positive, thus flatly disagreeing with both the preceding investigators. Hüttner,[5] using intervals from .1 sec to 3.0 sec says he found no indifference point. Kastenholz,[6] however, says Hüttner must have been laboring under some misunderstanding of the term indifference point, because his data really show four indifference points. Kahnt[7] found two indifference points. For a very short middle range of intervals, extending only from .5 to .7 sec, the errors were positive, but both above and below this range they were negative. Benussi,[8] however, disagrees with Kahnt, since he claims there is no error at either .72 or .36 sec. Katz[9] gives data with

[2] Briefly and somewhat inadequately, it may be said that by a negative error is meant that the second of two equal intervals which are compared is judged longer (first therefore conceived to be underestimated) and that by a positive error is meant that the second is judged shorter (first therefore overestimated). An indifference-interval is one which is regarded as separating intervals showing positive errors from those showing negative errors, and is itself subject to neither error. By the method of reproduction, errors are called negative when the reproduction is too short (stimulus-interval conceived as underestimated) and positive when the reproductions are too long (stimulus-interval conceived as overestimated).

[3] *Philos. Stud.*, 1885, **2**, 37–66.

[4] *Arch. f. d. ges. Psychol.*, 1924, **49**, 363–378. Most of this study was conducted with continuous stimuli, moving or stationary.

[5] Martius, G., *Beiträge zur Psychol. u. Phil.*; Bd. I, 1902, 367–410.

[6] *Arch. f. d. ges. Psychol.*, 1922, **43**, 171–223. The work of Kastenholz and others on the comparison of two empty intervals formed by three successive sounds is not here considered. This procedure has been strongly criticized by Benussi because it leads to pronounced and uncontrollable accent-rhythms or groupings.

[7] *Psychol. Stud.*, 1914, 279–320.

[8] *Psychologie der Zeitauffassung*, 1913, pp. 175–185. Actually, Benussi's data show a large negative error at 1440 σ, and an insignificant positive error at .36 σ (due entirely to the trials in which the variable was shorter than the standard). The data for .72 sec are said to be ambiguous and are not given.

[9] *Zeitschrift f. Psychol.*, 1906, **42**, pp. 327 and 339. Katz's data do not establish an indifference interval. They bear primarily upon the effect of the duration of the pause between the compared intervals.

two subjects which are not out of harmony with the hypothesis that the indifference interval has a length of .6 sec. Meumann,[10] who gives data for only one subject, with intervals extending from .1 to 1.0 sec, found an indifference point at .55 sec, errors being positive below this length and negative above it. Höring,[11] in a dissertation published in 1864 and inaccessible to the writer, is said to conclude that the indifference interval has a length between .365 and .454 sec.

The only justifiable conclusion from the preceding data is either that there is no definite, relatively constant indifference point or else that it has not yet been determined. How is it, then, that a common text-book statement about short time intervals is that there is an indifference point at about .7 sec? Such a statement does not seem adequately to summarize the literature just reviewed since no one of the investigators mentioned came to that conclusion or obtained data permitting such a conclusion. Well, there has been one other contribution to the question. Kollert,[12] in 1885, using intervals extending up to 1.2 sec, found an indifference point at .75 sec: but to arrive at this conclusion he was forced to discard something over ¼ of his data on no other grounds than that they were anomalous or due to contrast, the only contrast of which there was any evidence being that between the discarded data and those which were retained. Surely a result obtained by such means should not be given unquestioned preference over the findings of eight other investigators.

If the common statement of an indifference point at .7 sec rests upon data obtained in the case of the comparison of continuously filled intervals or in that of the reproduction of intervals, there is still less to justify it. With the method of reproduction, indifference points have been obtained as follows: Vierordt,[13] about 3¼ sec; Glass,[14] about 2⅜ sec; and

[10] *Philos. Stud.*, 1894, **9**, p. 282, Subject Rky. With another subject, but no pause between the intervals, Meumann found both .5 and .8 sec to be indifference points.

[11] Dissertation, Tübingen, 1864. Reviewed by Nichols, *Amer. J. Psychol.*, 1890, **3**, p. 504.

[12] *Philos. Stud.*, 1884, **1**, 78–90.

[13] *Der Zeitsinn nach Versuchen*, Tübingen, 1868, Pt. II. Experiments like those reported in Pt. I, in which the reproduction-interval is initiated by the terminal bounding sound of the stimulus-interval, so that the subject taps simply the terminal sound of the reproduction-interval, are not here considered.

[14] *Philos. Stud.*, 1885, **4**, 423–456.

Mehner,[15] two, errors changing from positive to negative at .75 sec but changing again from negative back to positive at 2.8 sec. H. C. Stevens[16] also found two indifference points, each obtained by a different method. With several reproductions of the stimulus-interval at each presentation, he found that all intervals up to .7 sec (the longest interval used) were overestimated. This result is the reverse of that obtained by L. T. Stevens,[17] who, likewise using multiple reproductions, found all intervals below .7 sec to be underestimated. By what he regards as the more reliable method, that of single reproductions, H. C. Stevens found an indifference interval somewhere between 2.4 and 3.7 sec, with intervals below this length *under*estimated and those beyond it *over*estimated. Schumann[18] used only intervals under .7 sec and found no indifference point. He contradicts Mehner and H. C. Stevens, however, in that he found the errors in this region to be negative instead of positive. Nichols[19] obtained results indicating an indifference interval in the neighborhood of .81 sec, with the shorter intervals underestimated and the longer ones overestimated. The subjects of Shaw and Wrinch[20] showed an indifference interval at from .175 to .250 sec. Lastly in the case of continuously filled intervals, we have the indifference point found by Quasebarth,[21] somewhere between 2–4 sec, and that of Hülser,[22] at about 1.8 sec. Kastenholz[23] found no general indifference point. Wrinch[24] found only positive errors below 1.5 sec and Thorkelson[25] found predominantly

[15] *Philos. Stud.*, **2**, 1885, p. 594. Most of Mehner's work was conducted by the method of comparison. With the method of reproduction, like Glass, he used only himself as subject. For a discussion of both, see The reproduction of temporal intervals, this JOURNAL, 1930, 13, p. 497.

[16] *Amer. J. Psychol.*, 1902, **13**, 1–27.

[17] *Mind*, 1886, **2**, 393–404.

[18] *Zeitschrift f. Psychol. u. Physiol. d. Sinnes.*, 1893, **4**, p. 63.

[19] *Amer. J. Psychol.*, 1891, **4**, p. 483.

[20] Univ. Toronto Studies, Psychol. Series, 1900, **1**, p. 144.

[21] *Arch. f. d. ges. Psychol.*, 1924, **49**, 379–432.

[22] *Arch. f. d. ges. Psychol.*, 1924, **49**, 363–378.

[23] *Arch. f. d. ges. Psychol.*, 1922, **43**, 171–228.

[24] *Philos. Stud.*, 1903, **18**, 274–327.

[25] *Undersøgelse af Tidssansen* (*Danish*), Christiana, 1885 (Reviewed by Meumann, *Philos. Stud.*, 1893, **8**, pp. 432–441).

negative ones above this point. The data of these last mentioned two investigators put together, then, suggest an indifference interval at 1.5 sec, though Thorkelson himself concluded that whether the errors turn out negative or positive is as much a matter of chance as whether a roulette wheel will come to a stop at the black or at the red.

In addition to the preceding studies are two by the writer,[26] conducted by the method of reproduction, the results of which suggest an explanation of much of the existing discord. In each of these two investigations, a group of subjects was taken through a set of intervals in regular order, from the shortest interval to the longest and then back again from the longest to the shortest. These investigations, in spite of a carefully planned and executed system of procedure, accurate control of stimulus-intervals and accurate measurement of reproductions, abundance of data, and careful and uniform instructions, certainly did not permit of the establishment of an indifference interval. Rather, on the whole, the positive and negative mean errors, or C.E.'s, were distributed in a manner not greatly different from what might be expected if their distribution were simply a matter of chance. One conclusion that seemed inescapable in view of the wide differences in the subjects' C.E.'s for the same interval, was that the subjects, while all uniformly instructed, were yet carrying out their instructions in different ways. This conclusion was confirmed through the differential instruction method, that is, through the use of different sets of instructions with the same subjects, and the determination of the effect upon the C.E.'s of this difference in instruction. Two sets of instructions were devised, which, although both requiring the subject to reproduce the stimulus-intervals as accurately as possible, gave results which showed a very large and significant difference in the C.E.'s.[27] Moreover, with one set of instructions the C.E.'s were predominantly positive and with the other set of instruc-

[26] The Reproduction of Temporal Intervals, this JOURNAL, 1930, 13, 473–499, and Individual Differences in the Reproduction of Temporal Intervals, *Amer. J. Psychol.*, 1933, **45**, 271–281.

[27] Individual differences in the reproduction of temporal intervals, *loc. cit.* See also the writer's article entitled Quotidian variability, *Psych. Rev.*, 1932, **39**, pp. 254–257.

tions, they were predominantly negative. The general conclusion, then, reached at this point, was that whether an interval is overestimated or underestimated, that is, whether the C.E. turns out positive or negative, depends in large part upon the subject, rather than upon the length of the interval, and in particular upon the internal processes of the subject, that is, upon the particular way he goes about the execution of his task.

This conclusion is believed to furnish the key to much of the contradiction between preceding investigators. With some subjects, one may obtain overestimation of any interval; with others, underestimation of any interval; and with still other subjects, one may obtain an indifference interval of almost any length. Yet most of the preceding investigators have used very few subjects, not infrequently only one subject (and that one has in several cases been the experimenter himself). In other cases, when the subjects have disagreed, the results obtained with the minority have been rejected. The brief instructions ordinarily used permit variation in the attitudes of the subjects and this variation is sufficient largely to account for the variety in the data obtained by different investigators. Undoubtedly many other factors enter into the explanation of the discrepancies found: variations in the order of sequence and in the range of intervals presented the subject; numerous differences in conditions and in methods; use of apparatus of various degrees of imperfection; and in many cases failure to maintain proper scientific standards of reliability. Many conclusions have been stated which were not reliably demonstrated to hold for even the few subjects used, and in other cases the data are presented in such a manner that it is impossible to determine the reliability of the author's conclusions. All these matters are involved in the explanation of the existing maze of discrepant findings; but certainly there is no more obvious and yet no more neglected factor than that of individual differences.

This conclusion that the type of error obtained depends upon the individual, does not, however, exclude the possibility that it may also depend in part upon the length of the interval.

But while the previous studies do not eliminate the possible existence of an indifference interval which is simply obscured by marked individual differences, they very definitely force the adoption of a statistical concept of this interval. They show that there is no interval which can be said in the case of all subjects exactly to separate shorter uniformly overestimated intervals from longer uniformly underestimated ones. They indicate that there can be no such thing as an indifference interval unless this interval be defined statistically in terms of some sort of an average or most probable outcome. At least three precise definitions, which differ slightly from each other, may be formulated. These will be stated and data will be presented which lead to a determination of the indifference interval in accordance with each of them. Since each of the three definitions refers to an average, it is obvious that the exact values indicated depend upon whether by the average is meant the arithmetical mean or the median.

First, the indifference interval may be defined as the interval at which 50 percent of the subjects give positive errors and 50 percent give negative errors. This definition is open to two objections, of which the more serious is that it leaves out of account the magnitude of the average errors. The mere sign of the errors is not an all-important matter. Two subjects showing errors of opposite sign may, if the errors are small, differ from each other far less significantly than two subjects showing errors of the same sign but of widely different magnitude. A further objection, of a practical nature, is that percentages near fifty are rather unreliable unless based upon a very large number of subjects.

Second, the indifference interval may be defined as the interval at which the average value of the average errors of the individual subjects is zero. As the average one may choose either the mean or the median error. The distribution of the reproductions, in the case of the majority of individuals, is skewed, with the median error lying to the negative side of the mean error, so that the median error is a larger negative or a smaller positive error than the mean error. As a consequence, the length of the indifference interval depends upon

whether it is defined as the interval at which the average of the mean errors is zero or the interval at which the average of the median errors is zero. This definition of the indifference interval, as the interval for which the average of the individual subjects' average errors is zero, is more satisfactory than one in terms of the percentage of subjects showing positive or negative errors, in that it takes account of the magnitude of the errors.

A third definition leads to the same length for the indifference interval as the preceding definition, but gives to its determination a far greater reliability. The errors of different individuals may be regarded as samples drawn from different regions of the more wide-spread distribution of all the errors of a large group of subjects;[28] and the real problem may be conceived as that of determining for what length of interval the mean of this larger, wide-spread, total distribution is zero. The mean of the total distribution, however, is determined through two selections or samplings, first, a sampling of individuals and second, a sampling (of fifty) of the errors of each individual. The only way in which these two samplings may be combined into one homogeneous kind is to ignore the individual subject and to substitute a vast population of internal processes (or of the corresponding errors), samples of which may be drawn from one individual. One may hesitate to ignore the tradition that psychology deals always with individual organisms; yet it may well be questioned whether science, at least when quantitative, ever deals with concrete wholes, except in so far as the whole is regarded in abstraction from the parts. Science deals primarily with abstractions such as the abstract attributes of length, weight, energy and intelligence, and with features or parts, such as the circulatory system, the nervous system, judgments, emotions and, in the present instance, internal attitudes and mean errors. Any definition of the indifference interval in statistical terms is certainly an abstraction; and there is no reason why this process of abstraction must halt at the concept of an average of

[28] In the present study, the P.E.$_{di.}$ of all errors and the average individual P.E.$_{di.}$ were, for the 600 σ interval, 45 σ and 38 σ, respectively, and for the 700 σ interval, 54 σ and 45 σ.

the individual subjects' errors. Instead of regarding the indifference interval as that at which the average of the individual subjects' average errors is zero, it may be taken as the interval at which the average of all the errors in any specified total distribution is zero. An illustration of one of these total distributions is given in Fig. 3, which represents the 5000 errors obtained from 100 subjects in the case of the 700 σ interval. This distribution is smoother and more regular than the distribution curve for simply the 100 individual C.E.'s (Fig. 2). The means of the two distributions are of course identical, but the reliability of the mean for the 5000 errors ($P.E._{dis}/\sqrt{5000}$) is much greater than that of the mean of the 100 individual C.E.'s ($P.E._{dis}/\sqrt{100}$). The reliability of the mean of the 100 C.E.'s is not greatly affected by the number of trials on which each C.E. is based, whereas the reliability of the mean of the total distribution takes into account the fact that each C.E. is based on fifty trials.

The indifference interval was determined in accordance with the statistical definitions given above. The apparently hopeless task of eliminating individual differences through a rigid control of the processes occurring within the subjects was not attempted. These internal processes were allowed to vary in whatever way they normally would, with the subjects instructed only to reproduce each interval as accurately as possible. If any general tendencies for the average errors to vary with the length of the interval existed, it was expected that these would be revealed by the use of sufficiently large numbers of subjects and of reproductions. Exactly how many subjects would finally be used was not decided at the start. It was planned to begin with four groups of fifty subjects (university undergraduates), each group to be given sittings with one of the four following intervals: 300; 600; 1200; and 4000 σ. After twenty-five students had been used with each of the intervals and a start made upon the second twenty-five with the interval 4000 σ (so that a total of thirty had been tested with this interval), a survey of the results obtained with groups of twenty-five indicated that it would be unnecessary to employ larger groups at the longer and shorter of the

intervals, because the errors were so one-sidedly positive at 300 σ and one-sidedly negative at 1200 and 4000 σ as to leave no doubt that the indifference interval was between 300 σ and 1200 σ. It was decided, therefore, that a more effective utilization of time and subjects would be obtained by increasing the number of subjects tested at the 600 σ interval to 100, since at this interval it was evident that the tendencies toward positive and negative errors were not far from equal. Finally, as a check on the results obtained with the 600 σ interval, it was decided to use another group of 100 subjects with an interval of 700 σ. The final data, then, consist of 50 reproductions of one interval by each of 280 subjects, divided with respect to the interval reproduced as follows: 300 σ, 25; 600 σ, 100; 700 σ, 100; 1200 σ, 25; and 4000 σ, 30.

Before the presentation of the obtained data, a brief description must be given of the conditions under which the experiment was conducted. The apparatus employed was with slight modifications the same as already described in detail in this journal.[29] The distinguishing feature of the apparatus was that it enabled the experimenter to measure the stimulus-intervals and the subjects' reproductions by means of a Hipp's chronoscope. Had this chronoscopic technique not been devised as a substitute for the older graphic techniques, the labor involved in the research would have been almost prohibitive. The set-up was such that if the chronoscopic measurement of the stimulus-interval was affected by any constant error, exactly the same constant error would enter into the measurement of the reproduction-intervals, so that the true error could be relied upon as the difference between the obtained measurements of the stimulus-interval and the obtained measurements of the reproductions. The stimulus-interval used at any sitting was always measured ten times immediately before and ten times immediately after the sitting, with the chronoscope controlled for the stimulus-interval used.

[29] *The Reproduction of Empty Intervals*, 1930, Vol. 13, pp. 478–482.

One modification in the apparatus previously described consisted in the use of a combination made by a local mechanic of a Meumann's time-sense machine and a General Radio synchronous motor [30] as the sole instrument for the control of the stimulus-intervals, in place of the Bergstrom large model chronoscope and the Bowditch-Baltzar contact clock used in the writer's preceding investigations. Another change consisted in giving the "ready" signal automatically, by use of a special, broad, flat contact attached to the circular rim of the Meumann's machine at a point which resulted in the signal (the squeak of an 800 d.v. buzzer lasting for 15 σ) sounding always 2.0 sec before the initial bounding sound of the stimulus-interval. This change made it possible to dispense with the pendulum previously used to enable the experimenter, by a key-pressing movement made in synchronization with the pendulum, to control the time of occurrence of the ready signal. One other change consisted in the use of a Klopsteg fall-control apparatus as a break-make key in the case of the short-interval 300 σ. To explain this change it is necessary to point out that if the subject's first tap is to start the chronoscope and his second tap is to do the opposite, that is, to stop it, some shift in electrical connections must be made between the taps. In the preceding work this shift was made by a break-make key operated by hand. With very short intervals, however, the experimenter could not always react quickly enough; so that with the interval 200 σ, it was necessary to dispense with the chronoscopic method and use the old, cumbersome, graphic technique of records taken alongside a tuning-fork on smoked paper. Even an interval of 300 σ is not long enough to permit with certainty the operation of a break-make key. It is for this reason that the Klopsteg, in suitable connection with several relays, was used with this interval. In the circuit closed by the subject's first tap was placed a relay so connected with the Klopsteg that as the current through the magnets of the relay was closed by the subject's tap, a circuit controlled by the relay and passing through the split-core magnet of the Klopsteg was broken. Thus almost immediately after the subject's first tap the steel ball held in place by the Klopsteg magnet began to fall. The position of this magnet was adjusted so that 150 σ later, the lower spring-contact of the Klopsteg was broken by the falling ball. This latter break resulted in a break in a circuit passing through two additional relays; but the contacts controlled by these two relays were so placed that as the arms of the two relays dropped forward away from the relay-magnets, a break occurred at one relay and a make at the other. The two relays together, therefore, exerted the action of a single break-make key; and this action occurred automatically 150 σ after the first tap of the subject. With the longer intervals a break-make key operated by hand was retained, since it obviated the necessity of resetting the Klopsteg fall-apparatus between trials.

The subject's taps operated the same sound-producing relay used in sounding the stimulus-interval. The storage-battery current actuating this relay was kept constant and, so far as could be detected, the sound remained uniform in intensity throughout the investigation. It is rather difficult to measure the intensity of an instantaneous, impact-sound such as the one employed. A determination was made, however, of the strength of an alternating current required to produce a tone just sufficient to cause the impact-sound to become unnoticeable, when the alternating current was run through a Baldwin, type C, telephone receiver. Two observers, the writer and Mr. G. R. Finch, a graduate student with experience in sound-intensity measurements, agreed that it

[30] See The effect of rate of sequence upon the accuracy of synchronization, this JOURNAL, 1931, 12, pp. 360–363, for the manner in which this instrument was used to control the intervals and data on its constancy.

required the full strength of 110 volts in the 60 cycle alternating current through the receiver to render the impact-sound inaudible, with the receiver and the sound-relay alongside each other and the observer in the position taken by the subjects.

After the subject had read the instructions, he was asked to repeat them in substance and, in the unusual case that he had misunderstood them, he was set right by oral restatement of the instructions by the experimenter. During the measurement of the reproductions the subject was alone in the subject's room, which contained only the tapping key, the signal-buzzer and the sound-producing electro-magnetic relay. The latter was concealed from the subject by a screen.

The key on which the subjects tapped was a Scripture noiseless key, adjusted so that a weight of 135 gms placed on the finger-button was just sufficient to press the spring bearing the button down to the position of contact. No study has been made of the effect of the amount of pressure required to operate the tapping key; but it is quite possible that a key requiring a very high degree of pressure, by causing more strain during the reproduction, would cause some increase in the length of the reproduction-interval. The subjects all tapped with two fingers, the middle and index fingers, held close together, and, after having been shown the way to tap, were given one or two minutes of practice at tapping (each tap producing a sound through the connected electro-magnetic relay) until they stated that they had no trouble in controlling the key and that the sounds always came at the intended instant. They were given no fore-practice, however, in the reproduction of intervals.

The data obtained are summarized in two tables, one dealing with the mean errors and the other with the median errors. Table I brings out the relation between the mean

TABLE I

THE RELATION BETWEEN THE MEAN ERRORS (C.E.'s) AND THE LENGTH OF INTERVAL

Intervals, C.E.'s and P.E.'s in units of thousandths of a second.

The P.E. of each C.E. ($P.E._{dis}/\sqrt{n}$) or of each percentage ($67.45\sqrt{pq/n}$) is given immediately under the corresponding C.E. or percentage.

Total number of subjects, 280; total number of reproductions, 14,000.

Interval	Ave. of Subjects' Mean Errors		Subjects		
	Mean	Median	Number	% + C.E.'s	% − C.E.'s
300	+18.7 2.5	+20.5	25	84 4.9	16
600	+2.0 2.4	−0.1	100	50 3.4	50
700	−5.9 2.9	−9.7	100	37 3.2	63
1200	−25.0 7.3	−25.1	25	32 6.3	68
4000	−183.8 61.2	−223.6	30	27 5.5	73

errors and the length of the stimulus-interval. Since the distribution of the mean errors of the individual subjects in the case of some of the intervals is somewhat skewed, both the mean and median value of the individual means are given. The table shows a regular and constant change, in the average mean error and also in the percentage of subjects giving positive or negative means, with change in the length of the stimulus-interval. The shortest interval, 300 σ, is on the average made too long in reproduction and with this interval 84 percent of the subjects show positive C.E.'s (mean errors); while the longest interval, 4000 σ, is reproduced on the average as too short, and with it 73 percent of the subjects show negative C.E.'s. The change from a positive mean C.E. to a negative one occurs between 600 and 700 σ. By linear interpolation between these two intervals, the value obtained for the indifference interval is 625.3 σ. This value is that of the indifference interval defined as the interval at which the mean of the individual mean errors is zero. Its probable error is probably not greatly in excess of that of the mean of the C.E.'s with the 600 σ interval, namely, 2.4 σ. While the reliability of this determination is incomparably greater than any hitherto made of the indifference interval as here defined, the reliability of the means from which it is derived is not as high as might be desired. For example, the positive mean for the 600 σ interval exceeds zero by less than the probable error of the mean, so that the true value of the mean C.E. for this interval may be negative. Similarly, the true value of the mean C.E. for the 700 σ interval may be positive. This low reliability follows from the fact discussed above that the obtained mean C.E. is the unweighted mean of only 100 numbers, each representing the C.E. of a single subject. It does not take into account the fact that each C.E. is the result of 50 measurements; so that the reliability of the mean C.E. as given in Table I is no greater than if the same C.E.'s had been obtained with single trials (instead of with 50). As already explained, the reliability of the mean of the total distribution of reproductions is much greater. If, then, the indifference interval be defined as that interval at which the

average of all reproductions, fifty from each subject, is zero, the present determination of the indifference interval immediately becomes a very reliable one. The probable error of the mean of the 5000 reproductions of the 600 σ interval is only .64 σ, and that of the 700 σ interval is only .74 σ. It

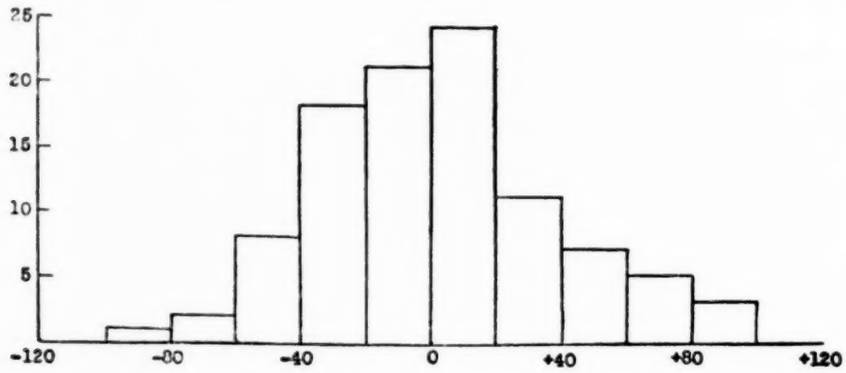


FIG. 1. Distribution of 100 mean errors with 600 σ interval. Mean = + 2.0 σ; P.E.$_{mean}$ = 2.4 σ; median = − 0.1 σ.

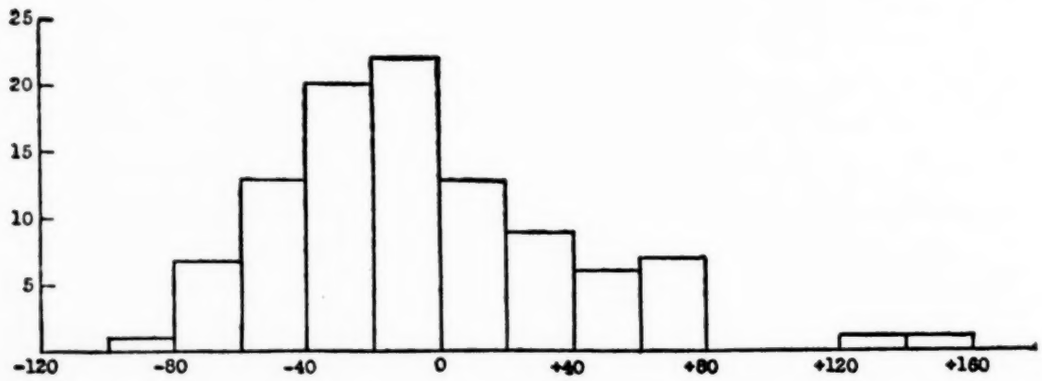


FIG. 2. Distribution of 100 mean errors with 700 σ interval. Mean = − 5.9 σ; P.E.$_{mean}$ = 2.9 σ; median = − 9.7 σ.

follows, then, that the mean of the reproductions of the 700 σ interval is reliably negative, negative to an amount equal to eight times its probable error. The mean of the 5000 reproductions of the 600 σ interval is less reliably positive, the small deviation from zero error being only three times its probable error. The indifference interval itself, when defined as the interval at which the average of all the errors is zero, must be very close to 625.3 σ, since the P.E. of each of the means from which this value is calculated, by linear interpolation, is less than three-fourths of one thousandth of a second.

If the medians of the individual means are used, the indifference interval is obviously very near 600 σ, since the median of the mean errors for this interval was negative only by one tenth of one sigma. By linear interpolation between 300 and 600 σ, and linear extrapolation from 700 and 600 σ, the values obtained are, respectively, 598.6 σ and 599.0 σ, the average of which is 598.8 σ. Since the probable errors of the medians are approximately 25 percent larger than those of the means, presumably the determination of the indifference interval from the means, at 625.3 σ, is, on account of its greater reliability, to be preferred.

The data on the percentages of subjects showing positive and negative C.E.'s are not considered to be as significant as those on the averages of the errors. It may be noted, however, that the percentage of subjects showing negative errors with the 700 σ interval, namely 63, is in excess of 50 by four times the P.E. of the percentage. The indifference interval, defined as the interval yielding as many positive as negative

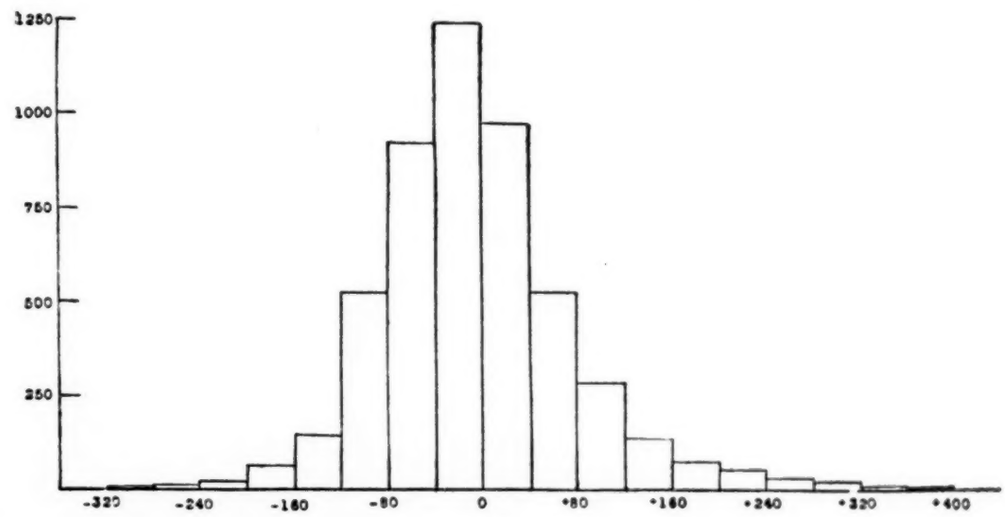


FIG. 3. Distribution of 5000 errors of reproduction with the 700 σ interval. Mean $= -5.9\,\sigma$; $\text{P.E.}_{mean} = 0.74\,\sigma$; median $= -13.5\,\sigma$.

individual C.E.'s, is exactly 600.0 σ, since with the 600 σ interval the division of subjects between the two types of C.E.'s was exactly half and half.

On account of the skewed distribution of the reproductions it is desirable to indicate the values obtained for the indifference interval when this interval is defined in terms of medians

instead of means. Due to skewness, the median error is more in the negative direction, or less in the positive direction, than is the mean error.[31] This skewness is evidenced by the fact that the median error was in the case of the majority of the subjects, with all intervals except the 1200 σ interval, more in the direction of negative error (or less in the direction of positive error) than was the mean error. The percentage of the individual subjects showing a median error more negative or less positive than the mean error, was, for the various intervals, as follows: 300 σ, 56; 600 σ, 66; 700 σ, 61; 1200 σ, 48; and 4,000 σ, 53. The distribution even of the medians was skewed in the case of some of the intervals, notably in the case of the 700 σ interval, with which the subjects' individual median errors had a mean of $-9.8\ \sigma$ but a median of $-17.9\ \sigma$. The skewness of the total distribution of errors with the 700 σ interval is obvious from the graph of Fig. 3. The mean of the total distribution is $-5.9\ \sigma$, while the median (calculated from grouped data) is $-13.5\ \sigma$.

TABLE II

THE RELATION BETWEEN THE MEDIAN ERRORS AND THE LENGTH OF INTERVAL

Interval	Ave. of Subjects' Median Errors		Subjects		
	Mean	Median	Number	% + Medians	% − Medians
300	+17.5 2.5	+14.5	25	84 4.9	16
600	−0.8 2.5	−1.0	100	49 3.4	51
700	−9.8 2.9	−17.9	100	31 3.1	69
1200	−26.6 7.1	−17.5	25	32 6.3	68
4000	−187.6 63.2	−209.9	30	30 5.6	70

[31] Judgments obtained in the case of the comparison of intervals by the method of constant stimuli display a similar skewness. See: Behavior with respect to short temporal stimulus-forms, Part I, Some hitherto undescribed illusions in the estimation of short intervals of time, this JOURNAL, 1928, 11, Tables II, V, VI and VII. For a discussion of the bearing of these results on the phi-gamma hypothesis in psychophysics, see Experimental evidence in support of Professor Thurstone's criticism of the phi-gamma hypothesis, by M. N. Fritz, *J. Gen. Psychol.*, 1930, 4, 346–352.

If the indifference interval be defined as the interval at which the average of the subjects' median errors is zero, then its length according to the data of Table II, is slightly less than 600 σ. If by the average of the medians is meant their mean, the indifference interval has a length of approximately 589.0 σ; if by the average of the medians, is meant the median, the value obtained for the indifference interval is approximately 587.5 σ. Each of these approximate values are obtained by averaging the value obtained by linear interpolation between 300 and 600 σ with that obtained by linear extrapola- beyond 600 σ, using the data for the intervals 600 and 700 σ. By the same method, the length obtained for the indifference interval when defined as the interval at which one half of the subjects give negative and one half of them positive median errors, is 594.4 σ.

The exact length obtained for the indifference interval thus depends upon the precise definition of that interval. The various definitions and the values of the indifference interval to which they lead may be summarized as follows:

Definition of Indifference Interval	Obtained Length
Interval at which zero is the value of the:	
Mean of subjects' mean errors	625.3 σ
Median of subjects' mean errors	598.8 σ
Mean of subjects' median errors	589.0 σ
Median of subjects' median errors	587.5 σ
Interval at which one half of the subjects give positive	
Mean errors	600.0 σ
Median errors	594.4 σ

The definition leading to the determination of greatest reliability is that which regards the indifference interval as the one at which the mean of all the errors of all subjects, considered as one distribution, is zero. According to this definition the indifference interval has a length of 625.3 σ. If the length of the indifference interval be expressed in units as coarse as tenths of a second, then it is a safe statement that, according to any of the definitions of this interval indicated above, its length is six tenths of a second.

It should be emphasized that the indifference interval as

determined in this investigation is an interval which is independent of the range of stimulus-intervals employed. This follows of necessity from the fact that entirely different groups of subjects were used with each stimulus-interval, so that it was impossible for the errors made by any subject to be influenced by sittings with other intervals. This point is important because the hypothesis has been advanced by Hollingworth [32] that the indifference interval is a function of the range of intervals used in its determination and a result of what is termed the central tendency of judgment. According to this view an individual working with a certain range of intervals tends to perceive them as all of one category of length, therefore to overestimate the shorter ones and underestimate the longer ones. In substantiation of this view as applied to the temporal indifference interval, Hollingworth constructed a table of results from six investigators showing some correspondence between the indifference interval and the range of intervals used. Most of the investigations cited are decidedly unreliable and can hardly be regarded as affording decisive evidence. A more thorough and accurate review of all the investigations of the indifference point would show that the existing discrepancies cannot possibly be explained as due solely to a law of central tendency of judgment. Even a considerable portion of the data cited by Hollingworth in support of his hypothesis are opposed to it. For example, the subjects of Estel worked with intervals varying from 1.5 to 8.0 sec. The shorter of these intervals should, by Hollingworth's hypothesis, have been overestimated. But Estel found *all* of the intervals he worked with to be *under*estimated. That Estel noted a periodic fluctuation in the magnitude of this underestimation does not detract from the fact that at no point did he find overestimation. Again, L. T. Stevens, who is also among the group of investigators cited by Hollingworth, obtained results diametrically opposed to the law of central tendency, since he found the shorter intervals to be *under*estimated and the longer ones *over*estimated, whereas

[32] *Arch. of Psychol.*, 1909, 2, No.13, pp. 20–33 and *J. Philos., Psychol. and Soc. Meth.*, 1910, 7, 461 ff.

according to Hollingworth's hypothesis just the reverse should have been found. Some day, no doubt, a thorough and systematic investigation of the effect of reactions to intervals of various lengths upon those of other lengths will be published. Such an investigation is needed because these effects exerted by preceding stimulation upon present response constitute an interesting psychological topic and they are, perhaps, in no field more marked than in that dealing with short temporal intervals. It is probable that such an investigation would meet with difficulties since it is likely that the effects of work with one interval upon that with another interval are very complex. While such an investigation would no doubt afford abundant evidence of the phenomenon of assimilation and data which could be interpreted in terms of Hollingworth's law of central tendency, it would also afford information concerning the limits of this law, and in particular throw light on the question of when errors of assimilation change to errors of contrast; for of the phenomenon of contrast in time-perception there is as little doubt as of that of assimilation.

In the case of the reproduction of intervals, work with one interval may influence that with a different interval (or with the same interval at a different sitting) through factors which may be even more important than those of adaptation, assimilation and contrast. A very peculiar factor was revealed in the present study as a result of asking the subjects, at the end of the sittings, whether they thought that on the average they had reproduced the intervals as too long or as too short. Frequently the subjects seemed somewhat in doubt, though they often quite promptly gave the reply 'too short' or 'too long,' a reply which sometimes agreed and sometimes disagreed with the actual measurements. Now since they understood that they were to make the reproductions equal to the stimulus-interval, and tried hard to do so, it seemed somewhat surprising that they should confess that, in thinking the matter over afterwards, they did not feel that they had entirely succeeded. Whether or not judgments concerning their average errors would have occurred to the subjects if

they had not been questioned on the matter is not known. It seems plausible, however, that, had the subjects been used a second time, some of them would have approached the second sitting with the idea that they should make the reproductions either longer or shorter than at the last sitting, while others would have begun the second sitting without reflecting upon previous errors and therefore without intention to introduce any correction in their previous behavior or any modification in the way they should carry out the instructions. It might well be, then, that the tendency to correct or not to correct supposed errors at the previous sitting would be an important factor in determining performance at a second sitting, and a factor which might be of importance in accounting for individual differences in the reproduction of intervals when subjects are used at more than one sitting.

At first the questioning of subjects at the end of the sittings as to what they thought had been the direction of their errors was carried on as a matter of personal curiosity and was not regarded as part of the experiment, so that no record was kept of the subjects' replies. The curious fact soon appeared, however, that the subjects usually judged that on the average their reproductions were too short rather than too long. Thereupon answers were recorded from 67 consecutive subjects with the 600 σ interval and from 50 consecutive subjects with the 700 σ interval. It seemed likely that some subjects, simply in order to avoid creating the impression that they had not tried conscientiously to carry out the instructions, would say that their reproductions had been of about the correct length. Consequently the subjects were asked to *guess* whether on the average they thought their reproductions were too short or too long. In case the reply was that the reproductions seemed to be of the correct length, a guess was not insisted upon. If replies such as 'too short if anything' or "most of them were about right but some of them were too short" are recorded simply as 'too short' and similar replies using the word long instead of short as 'too long,' the results obtained may be indicated as follows:

Actual mean error	Positive			Negative		
Guessed mean error	Pos.	Neg.	Zero	Pos.	Neg.	Zero
Interval 600 σ	7	19	3	7	26	5
Interval 700 σ	2	12	1	5	24	6

This tabulation shows that about 70 percent of all replies were 'too short'; and this percentage was just as high when actually the reproductions were too long as when they were too short. The subjects tried to make the intervals the same, but afterwards they did not always seem so. One subject volunteered the comment that "The intervals seemed right when tapped, but afterwards too short. It seemed hard, for some reason [which the subject could not specify] to make them longer." Another subject said, "I tried to make the reproductions equal to the interval you gave me, but afterwards they sounded too short. Then I tried to make them longer but still they seemed too short."

This tendency on the part of a majority of the subjects to judge at the end of the sitting that their reproductions have been shorter than they should be suggests that the indifference interval obtained by the method of reproduction may be shorter than that which would be obtained by the method of comparison. It would be unsafe, however, to make predictions concerning the indifference interval in the case of the comparison of intervals from results obtained by the method of reproduction. The attitude of the subjects is different throughout in experiments conducted by the two methods. Decision concerning the relation of the reproduction indifference interval to other indifference intervals must be deferred until the latter have been more precisely defined and more dependably determined. The writer ventures the belief, however, on the basis of both published [33] and unpublished [34] data, that in the case of empty intervals, bounded

[33] Behavior with respect to short temporal stimulus-forms, Part I, this JOURNAL, 1928, 11, pp. 188–190.

[34] W. A. Blakely, doctor's thesis, 1932, Department of Psychology, Univ. of Illinois.

by instantaneous sounds, the comparison indifference interval is practically identical with the reproduction indifference interval. There appears also to be an absolute judgment indifference interval of the same length.[35] On the other hand, the few existing studies of the comparison of continuous durations are not sufficiently dependable to justify even a surmise as to whether the indifference interval for continuous durations is identical with that for empty intervals.

(Manuscript received February 23, 1933)

[35] For a tabulation by the writer of results of five investigations, see this JOURNAL, 1928, 11, p. 190.

SINGER'S DEFINITION AND THE GENERALIZED LAW OF THE GEOMETRIC MEAN IN NUMERICAL ESTIMATION

BY C. C. LIENAU
West Virginia University

Introduction

The fact that psychophysical statistics generally require skew frequency curves for their description has been given greater or less recognition since the time that Seidel employed the geometric mean in star photometry.[1] Galton and McAlister were apparently first to formulate explicitly [2] the Law of the Geometric Mean. This law has since engaged the attention of many investigators, some of whom came upon it independently.

Fechner used an algorithm for computations [3] which clearly implies it. Kapteyn's system of skew frequency curves [4] contains several generalizations of the law and he discussed at some length its application to data on the threshold of sensation. His conception of the transformation of normally distributed variables was criticized by Karl Pearson [5] who mentions experiments of his own on the comparison of tints and lengths. Hertzsprung [6] has shown the simpler law to be useful as a graduation function for König's data on the intensity limen of the eye. Shewhart [7] has applied the law to 710 observations on the threshold of audition with the unusual goodness of fit of 75 percent. More recently Thurstone [8] and Fritz [9] have shown that for coarse discrimination the Phi-Log-Gamma Hypothesis, which is the law of the geometric mean in integral form, is a better fitting psychometric function than the Phi-Gamma Hypothesis. Thurstone regards the use of this function as an empirical device.

The law of the geometric mean can be viewed as the result of transforming a normally distributed variable to another of which the first is the logarithm. If this transformation

function be generalized so as to make it independent both of units and zero of the second variable, say x, we obtain Kapteyn's first generalization, equation (3) below.

A transformation of this sort, *i.e.*, of the form of (2) below, is identical in form with E. A. Singer's Definition of Sensation-Intensity.[10] It will be seen to reduce to Fechner's when $x_0 = 0$ and GK' is taken as the threshold stimulus. That it should so reduce is more than coincidence for Singer's formula results from the application of function theory to a set of general postulates. For details of its development the reader is referred to Singer's original paper.

In this paper are discussed, in relation to Singer's definition, four experiments on a coarse type of equality judgment variously known as estimation, prizing, or guessing. A subject is required to estimate and specify the number of a group of objects. Many trials yield a frequency distribution of estimates with the set of all positive integers as the variable class. A study of the distributions for groups of plane and three-dimensional objects numbering 10 to 5000 showed that by formal criteria the generalized law of the geometric mean is a very probable description of such series and is otherwise a logical description of relevant facts.

Upon interpretation after the manner suggested by Kapteyn it soon appeared that Singer's definition is a valuable aid in taking the step from mere description to prediction. In fact, calculations from his fourth postulate enable one to predict with fair accuracy the smallest estimate likely to appear in a given series, given the relation between the limen and the standard stimulus magnitude. Overestimation, whether due to fatigue or illusion, is also conveniently and logically described by Singer's parameters.

In the succeeding paragraphs the discussion treats first the formal properties of the distribution law. A statement of Singer's definition and some of its implications are then given, followed by a discussion in which a picture of the first experiment and its results helps to make more concrete the relation of Singer's formula to the probability law in its present and analogous applications. The remaining experiments and

the irresults are then briefly described. Finally, a tentative relation between the limen and the standard stimulus is calculated from experiment for the purpose of applying Singer's fourth postulate.

Generalized Law of the Geometric Mean and Its Properties

Let a variable σ be distributed according to the normal law, with unit standard deviation and a mean value of zero. The chance that σ does not exceed a stated value is then

$$P(\sigma) = [\sqrt{2\pi}]^{-1} \int_{-\infty}^{\sigma} \exp\left\{ -\frac{1}{2}\sigma^2 \right\} d\sigma. \tag{1}$$

Let the relation between σ and x' be defined by

$$\left.\begin{aligned} \sigma_1\sigma &= \log_e\left(\frac{x' - x_0'}{G - x_0}\right)\left(\frac{1}{K'}\right) \\ x_0' &= K'x_0 \end{aligned}\right\} \quad \begin{aligned} &\sigma_1 > 0 \\ &x' > x_0' \\ &G > x_0 \\ &K' > 0 \end{aligned}, \tag{2}$$

where K', G, σ_1 and x_0' are independent of x'.

The chance that x' falls within the interval dx' when its relation to σ is given by (2) must equal the corresponding chance that σ falls within $d\sigma$. Making the transformation we obtain for the chance that x' will not exceed a stated value u:

$$P(u) = [\sqrt{2\pi}\sigma_1]^{-1} \int_{x_0'}^{u} \exp\left\{ -\frac{1}{2\sigma_1^2}\left(\log_e \frac{x' - x_0'}{(G - x_0)K'} \right)^2 \right\} \frac{dx'}{x' - x_0'}. \tag{3}$$

This law has 3 degrees of freedom, G, x_0, and σ_1, unless restraints are imposed upon these parameters.

If N' is the expected number of times that x' takes any value whatever, the expected number of instances in which x' is less than a stated value u is:

$$n' = N'P(u). \tag{4}$$

The integrand of (3) is represented by a skew probability curve limited at the left by x_0', rising to a single mode near GK' which is the most probable value of x', finally vanishing as

$x' \to \infty$. GK' is also the median, and the geometric mean taken about x_0'. Instances measured by x_1 and x_2 are equally likely when $x_1 < GK' < x_2$ are in geometric progression with respect to x_0'. The form due to Galton and McAlister is obtained by setting x_0' to zero. From their point of view the law applies whenever the geometric mean is the most probable value of a series; as, vital phenomena, where Fechner's 'Law' applies, etc. The present generalization is valuable in that the geometric mean is freed from accidental choice of origin and has in this a useful relation to Singer's fourth postulate.

To fit the differential curve to raw data Karl Pearson [5] has calculated the moment formulae for the case $x_0' = 0$ and these can readily be extended to include the present case. Davies [11] has given the Beta curve. The value of x_0' is infected with a relatively larger sampling error than σ_1 and G. When x_0' is not too large or can be estimated, the series can be reclassified on the σ scale to which the Normal formulae can then be applied.

Of greater interest and convenience for psychophysical data is the use of logarithmic probability paper to represent the integral formula. When $x_0' = 0$, $P(u)$ *vs.* $\log_b x'$ gives a straight line of slope proportional to $1/\sigma_1$. When $x_0' > 0$, $P(u)$ is asymptotic to x_0' for x small, and to the straight line of equal σ_1 and $x_0' = 0$, when x is large. When x_0' is negative the curvature is opposite, but in either case it may be estimated by rectifying the curve. Of especial interest as an example of Kapteyn's graphical method is the fact that the curve $P(u)$ also gives the relation between σ and u. The null point of σ is at $P(u) = 1/2$ and $\sigma = +1$ is approximately at $P(u) = 0.84131$ with other values in proportion. If we accept Singer's definition of this function in psychophysical applications of (3), equal distances along the probability scale measure equal sense differences.

Singer's Definition

(*a*) *General*

The failure of experimentalists to give their attention to Singer's formula is partly explained by the concise nature of

its development. Perhaps more troublesome is the uncertain relation of its parameters to the statistics of behavior. A recognition of the fact that each parameter can have a counterpart in (3) makes empirical interpretation seem the less a *tour de force*.

One may, however, temporarily divorce the discussion from probabilities, regarding the formula as a mean relationship. But it is precisely at this point that the question arises: What mean is representative? It has been shown that the postulate of the arithmetic mean as the most probable of a series is not always, though very frequently, appropriate. If the simple geometric mean be representative of vital phenomena as Galton claimed, then the law of the geometric mean is indicated. If this mean be generalized with respect to origin and postulated as the most probable of the series, then the generalized law (3) is indicated. But this may be analyzed into the Normal Law, and an expression identical in form with Singer's definition. Thus we see that the application of the definition and the proper treatment of relevant experimental material are not strictly separable problems.

Nevertheless, considering Singer's final result by itself: we have a quantity $\phi(r_1)$, which he termed the "difference of sensation-intensity" between an arbitrary stimulus r_1 and a reference stimulus r_ω measured in a scale of physical intensities, determined by the function:

$$\phi(r_1) = a \log_e \frac{r_1 - r_0}{r_\omega - r_0} \qquad a > 0, \quad \begin{cases} a > 0 \\ r_1 > r_0 \\ r_\omega > r_0 \end{cases} \tag{5}$$

where

$$a = \left[\log_e \frac{r_{1+} - r_0}{r_1 - r_0}\right]^{-1}, \quad r_4 > r_1 \tag{6}$$

$r_{1+} - r_1$ being the limen at r_1 and r_0 the 'physiological zero.'

The first underlying postulate asserts the self-consistency of the sense scale. The second states that any general sensation formula must be independent of such accidents as, for example, the non-coincidence of the null points of various temperature scales. This requires the presence of the term r_0.

The third restates the requirement of dimensional homogeneity so well understood by physicists, thus yielding with the second postulate the quadrad form for quantities involving r. The fourth states that four quantities, r_1, r_2, etc., are always in geometric progression with respect to some origin. This serves to define r_0 as the datum of a geometric progression in which the limen figures prominently, a relation to be treated in detail presently. The fifth makes r_ω the stimulus for which $\phi = 0$, *i.e.*, it fixes the datum of the ϕ continuum. The sixth assumes that under constant subjective conditions, continuity for ϕ and its derivatives is permissible within the probable error.*

Postulates 7 and 8 state that under constant subjective conditions the rate of change of ϕ with r_1 must be independent of the choice of datum in the ϕ scale, and that ϕ increases with r_1. The requirements so far are satisfied by (5). To evaluate the unit of the ϕ scale the next assumption fixes upon r_1 and its limen, giving (6) for (*a*), which will here be termed the sensibility. The sensibility is seen to be an increasing function of the density of probably discriminated stimuli in the neighborhood of r_1.

Finally, to interpret r_ω; the stimulus r' that under the conditions of overestimation is judged equal to r gives the tenth postulate: $\phi(r) - \phi(r') = 0$. The solution of this equation gives $r_\omega = r_1'$, provided r_0 and its analogue r_0' are not very different; and, for the sensibility in overestimation Singer obtains

$$\frac{\mathrm{I}}{a'} = \log_e \left(\frac{r_{1+} - r_0}{r_{1+}' - r_0'} \right) \left(\frac{r_1' - r_0'}{r_1 - r_0} \right). \qquad (7)$$

(*b*) *Physiological Zero*

Returning now to r_0; the fourth postulate states that this is the limiting solution of

* When the use of continuous functions, such as Fechner's definition, is disputed,[12] it is usually on the grounds of the "all or none" principle. The discrete nature of the microscopic physiological events involved does not, of course, prohibit us from approximating the probability of their occurrence, or average relations between macroscopic variables like r_{1+} by means of continuous functions. The fluctuation in any parameter of behavior, the limen for example, is generally large enough to permit this. As the limits of the probable error are narrowed it may eventually be necessary to posit discrete events.

$$\frac{r_{1+} - r_0}{r_1 - r_0} = \frac{r_{2+} - r_0}{r_2 - r_0} \qquad \text{as} \qquad r_2 \to r_1, \tag{8}$$

where r_1 and r_2 are arbitrary stimuli and $r_{1+} - r_1$, $r_{2+} - r_2$ define their upper limens.

The existence of this limit supposes that r_{1+} can be expressed as a continuous function of r_1 with at least one derivative. Let this be written

$$r_{1+} = f(r_1) \tag{9}$$

and the solution of (8) becomes

$$r_0 = \left(r_1 \frac{df}{dr_1} - f \right) \div \left(\frac{df}{dr_1} - 1 \right). \tag{10}$$

If no further restrictions are imposed on r_0, f and its derivative are unrestricted except that if r_0 is to remain finite the denominator must not vanish unless the numerator does so at the same time. For, let $df/dr_1 = 1$.
Then

$$r_{1+} - r_1 = \text{const.} \tag{11}$$

If the numerator does not vanish we have since $r_{1+} > r_1$

$$r_0 = -\infty. \tag{12}$$

That is, if we reject (12) the case of constant limen is excluded unless the numerator vanishes, *i.e.*,

$$r_{1+} = kr_1. \tag{13}$$

Since (12) and (13) cannot be true simultaneously unless the absolute limen is zero, we should reject the experimentally impossible case. This is (12), for zero limen implies infinite sensibility, *i.e.*, $k = 1$. This result has an important statistical consequence to be deduced later.

Let r_0 be constant. The solution of (10) gives

$$\frac{r_{1+} - r_0}{r_1 - r_0} = \text{const.}, \tag{14}$$

or Weber's Law holds with respect to a constant degree of the r scale. Like Weber's Law this expression is generally inadequate to the facts.

We need not proceed to the limit to define r_0. In (8) let $r_2 = r_{1-}$ and $r_{2+} = r_1$, and take r_0 as the solution of

$$\frac{r_{1+} - r_0}{r_1 - r_0} = \frac{r_1 - r_0}{r_{1-} - r_0}. \tag{15}$$

Therefore r_{1-}, r_1 and r_{1+} are in geometric progression with respect to a datum r_0. If they are not very different as in fine discrimination (10) and (15) give nearly the same result. The extreme stimuli might as shown below be interpreted as those equally likely to be discriminated from r_1. Since r_0 is redefined for every stimulus r_1, neither Weber's Law nor any other is implied by the truth of (15).

Experimental interpretations like the lower threshold, the null point of an arbitrary temperature scale relative to the thermodynamic, etc., can suggest limitations on r_0. Dr. Singer has suggested the weight of the arm or hand in lifting experiments. Measures of these quantities could be compared with r_0 calculated from the limen and its variation with r_1 to test (8) and its variants.

If r_0 can be shown to be the lower bound of a probability function like (3), its interpretation will become a relatively simple matter. Variation of r_1 will furnish related data on the limen. Calculations from (8), etc., can then be compared with the estimated lower bound of the statistical series.

Experimental Probabilities

Statistical questions are raised as soon as we turn to the empirical meaning of r_{1+}, r_{1-}, etc. To answer some of these more definitely we can refer with advantage to the conditions and results of the present experiments.

It is only fair to state that these experiments were not from the outset designed for the purpose of applying the definition. The only preconception in that direction was that the frequency law of the estimates might in some way exhibit the operation of Weber's Law. The subsequent failure of Weber's Law and the fact that the generalized law of the geometric mean which introduced an adjustable lower bound, was also capable of describing a tendency to overestimate, led

through Kapteyn's method of interpretation directly to Singer's formula. In a new series of experiments, undoubtedly worth undertaking, the technique should be open to less criticism than that of the present set which were performed in a rather casual manner.

Experiment I gave a distribution in striking accord with (3), the more so as the data were taken from a 'guessing contest' not prearranged to prove any hypothesis. Not greatly complicated by unexplained factors, it makes a suitable basis for discussion.

One hundred ninety-eight incandescent lamps of the familiar opal glass, flask-shaped type, $2\frac{1}{2}'' \times 6''$, were placed in a roughly conical heap on view in the window of a shop on the main business street of a town some 20,000 in population. The pile was approximately 30″ in diameter and 18″ high. It could be viewed from various points within an angle some 120° in a horizontal plane and was roughly 3′ below the average observer's eye. A sign reading: "Can you guess the number of lamps in this window?" gave invitation to enter the shop to record the guess. Small prizes were offered for the closest estimates. The configuration of the display was not changed during the contest which lasted 14 days. The window was generally illuminated on evenings when the shop was closed.

Each estimate was recorded under the last by the subject, who was neither permitted to see those previously made until his own had been entered, nor to enter more than one. On occasion two or more persons entered the shop together, possibly having reached a joint decision. The clerk who kept the record remarked upon the fact that a judgment once passed left that subject reluctant to change it, even after seeing the record. No special effort was made to assure independence of successive estimates except as noted above. The contest did not attract a very large number of persons. The resulting 136 estimates appear in an appendix with temporal order preserved.

These estimates are distributed in close agreement with (3) about a mode near 198. The median and simple geometric

mean both equal 191. A value of $K' = 0.97 \pm 0.10$ does not conclusively point to underestimation. The value of x_0 is very probably zero. The extreme estimates 60 and 643 have a simple geometric mean 196 and an arithmetic mean 352. The results of applying Pearson's method are given in table (1). If the contest had been terminated with the 68th

TABLE I

EXPERIMENT I

Estimates,	$N' = 136$
Probable Lower Bound,	$r_0' = 0$
'True' Count,	$r_1 = 198$
Error 'True' Count,	$= 0$
For ungrouped estimates,	$\log_{10} K'r_1 = 2.2806$
For ungrouped estimates,	$\log_{10}$, Median $= 2.2810$
Median,	$= 191$
Geometric Mean,	$= 191 \pm 19$
Estimation Factor,	$K' = 0.965 \pm .10$
	$\sigma_1/2.30 = 0.1569$
Standard Error of $\text{Log}_{10} K'r_1$	$= 0.0135$
Standard Error of σ_1 ($\div$ 2.30)	$= 0.0095$
Asymmetry,	$\sqrt{\beta_1} \pm \sqrt{6/N'} = 0.0970 \pm .2100$ For Normality $= 0$
Flatness,	$\beta_2 \pm \sqrt{24/N'} = 3.2355 \pm .4204$ For Normality $= 3$
Type Criterion,	$3\beta_1 - 2\beta_2 + 6 \pm \sqrt{96/N'} = 0.4425 \pm .8205$ For Normality $= 0$

Class Limits		Actual Frequency f	Expected Frequency f_e	d*	Expected† 1A and 1B	Actual 1A	d_a	Actual 1B	d_b
∞	1.9528	2	2.495	0.100	1.25	2	0.40	0	1.25
1.9528	2.0558	8	7.83	0.037	3.91	5	0.31	3	0.21
2.0558	2.1588	19	19.42	0.009	9.71	9	0.05	10	0.01
2.1588	2.2618	32	31.76	0.002	15.88	16	0.00	16	0.00
2.2618	2.3648	37	34.26	0.219	17.13	17	0.00	20	0.49
2.3648	2.4678	24	24.39	0.006	12.20	13	0.05	11	0.12
2.4678	2.5708	11	11.49	0.021	5.75	6	0.01	5	0.09
2.5708	∞	3	4.38	0.435	2.19	0	2.19	3	0.30
Totals..		136	136.02	0.829	68.01	68	3.01	68	2.47

Goodness of Fit, P = .97 P = .85 P = .87

* $d = \dfrac{(f - f_e)^2}{f_e}$.

† From the same equation as for f_e.

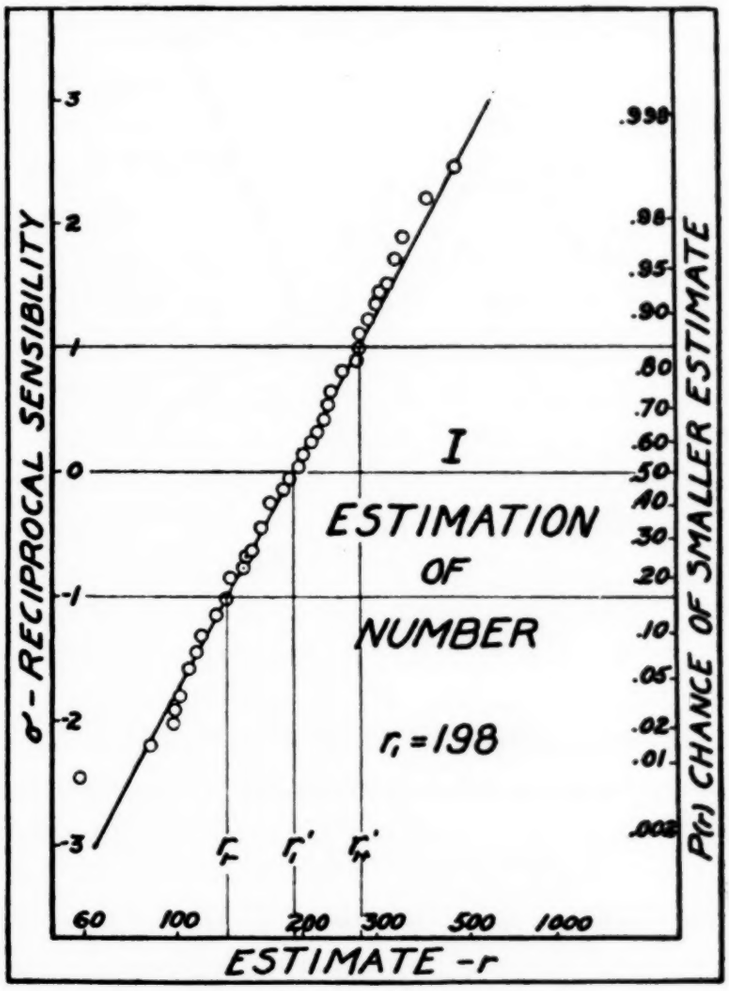
estimate, a curve fitting the first set should have fitted a second set of 68 equally well within the sampling error.* The

* This should be distinguished from the error made by the experimenter in classifying the stimulus. In this case the probable error in counting 198 lamps is negligible. The error in Urban's weights is of the order 0.1%. The sampling error of a psycho-

way in which this supposition was tested can be seen by inspection of table (1) and figures (1) and (2).

Fluctuations in the course of the summation for small deviations from the median are of greater weight than those in the

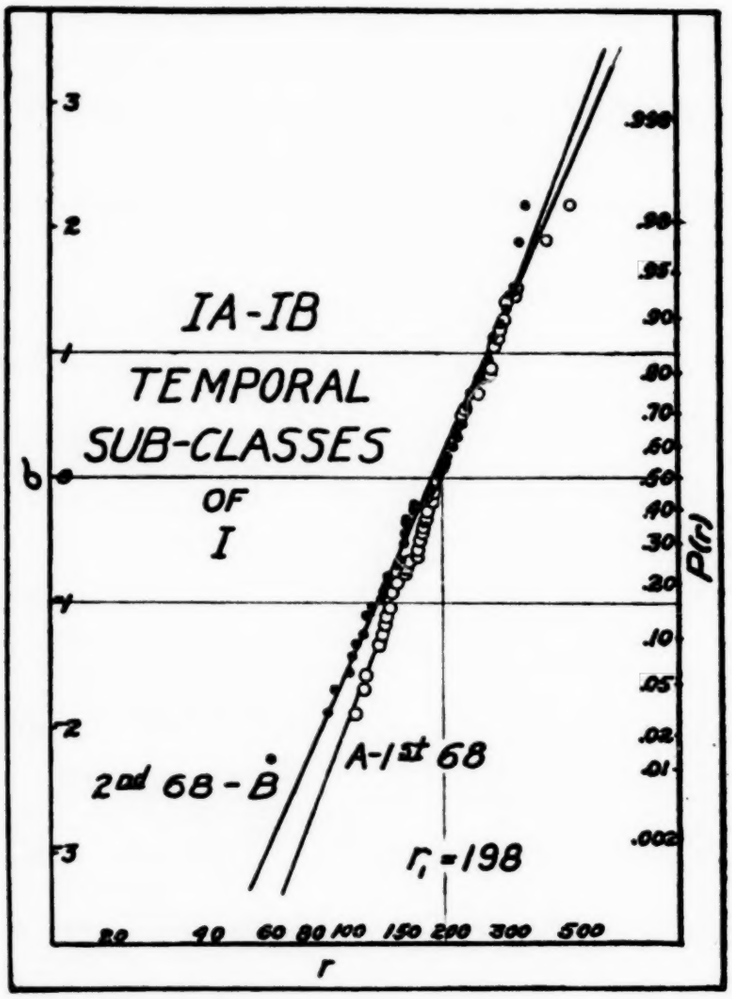
FIG. 1



tail classes since the latter are infected with a relatively larger sampling error. The value of x_0, if larger than zero, appears to be too small to stand out from the sampling fluctuations in that range. This consistency with the main features of the law of the geometric mean makes it a very probable description of such series.

physical parameter, dependent as it is upon the standard deviation of the psychometric function and the number of judgments, is therefore a character of the 'universe' of subjects. The resultant 'probable error' of the parameter is compounded of the experimenter's confusion of classes and fluctuations in the behavior of his subjects.

FIG. 2



THEORETICAL PROBABILITIES

(a) *Variable Class*

The judgment involved here consists in this: under fixed conditions a group of objects is set by estimation in correspondence with one of the set of classes denoted by $\cdots -2, -1, 0, 1, 2, \cdots, \cdots m, m+1, \cdots$. Each member of this set is greater than the preceding and can be derived from it by the operation $m + 1$, where m is any integer or zero. If the group be composed of physical objects the operation $1 - 1 = 0$ has, but $0 - 1$ has no physical counterpart. Hence, estimates as to the number which the physical operation of counting will yield must be limited to $0, 1, 2 \cdots m, m + 1, \cdots$.

Let the result of one set of conditions (a direct count, say) be the judgment m_1. Let another set of conditions for the same group lead to the judgment m'. Then m' will be equal to or different from m_1. According as m' or m_1 be taken for the datum, the second judgment is an over or under estimation of the stimulus with respect to the first judgment. The choice is pure convention but may be made more reasonable from considerations of probable reliability, utility, etc. Let m_1 be the preferred datum.

Let there now be made N' trials under the same conditions by as many different subjects leading to N' instances of m'. Experiment shows that while any m′ may appear more than once the estimates do not as a group agree, the judgments being more or less widely dispersed with certain proportions falling on the average in each class. An a priori law must provide a probability for all and only those classes m' into which the judgment can fall under the conditions of the experiment.

As one of these conditions it can be stated in advance that the lower bound is equal to or greater than zero. More than this cannot well be done. For, while the infinity of variables determining the final class of the judgment can be dealt with symbolically, there is no assurance that relations postulated are realized, unless we are content with certain broad assumptions as to their nature. If the many variables are also small in their individual contribution to the final result, a second supposition of the Normal Law of error is realized. But an outright assumption of normality for m' violates the requirement that negative classes be excluded, for the normal curve extends to $-\infty$. This circumstance results from the assumption of the arithmetic mean of the series as its most probable value. Then what sort of mean is applicable here? The general geometric mean will be postulated, but rather than proceed at once to the deduction of the corresponding probability law we shall digress somewhat from systematic argument in order to find an empirical meaning for Singer's parameters.

In the present experiments the apparent physical absence

of the stimulus with which the group of objects is compared is rather puzzling unless we recall that another aspect of the group is the m class. All instances of the sets m up to m_1 are thus actually present, if not visibly then to inference at least. But while any one of these classes might serve as comparison stimulus, classes $m > m_1$ are physically absent. This sort of difficulty has led some to postulate an absolute scale resulting from experience.[13] Some degree of discriminal ability is implied by every estimate since subjects incapable of such judgment are excluded by definition of behavior along with those who do not express an opinion. Consequently, such a scale may be a helpful concept. But it evidently corresponds to no simple reality, and any complex one we can imagine may leave out the very aspect of the stimulus on which some subject bases his judgment. It seems preferable to let experiment define an abstract self-consistent sense scale whose unit is measured in terms of the confusion of stimuli in the physical scale.

(*b*) *Physical Scale and Overestimation*

Singer classes physical dimensions of a stimulus as intensive or qualitative, the latter being its non-energy aspects. Such would be the pitch of a pure tone. It is not intensive and is an independent dimension of the stimulus. A given stimulus has many such dimensions. The limitation of the formula to intensive scales is not necessary. It is one purpose of this paper to show that the identification of r_1 with that obviously qualitative aspect of a group, its number, leads to a useful interpretation for all parameters of (5), (6) and (7).

That not all aspects of a group are in simple correspondence with number is clear. A uniform stack of equal cubes occupies a volume of space in one-one correspondence with the number of cubes until uniformity is destroyed. The presence of the interstices requires the introduction of additional dimensions before the total space occupied by the group can be specified. The same is true for other configurational group properties, for light emitted or reflected, its quality, etc. Let there be $v + 1$ such dimensions and e equations of restraint

defining a stimulus. All such stimuli are in one-one correspondence with points in $(v + 1 - e)$ space.

A random estimate of number may be based on any or all these aspects. If one set of such aspects predominates in effect, bias of the mean judgment will probably result. Ponzo has shown geometrical group properties to be correlated with a tendency to over or under estimate number.[14] Another example is given by the trade trick of supporting a relatively small quantity of candy with false work to create the impression of a shop window lavishly filled. In experiment II was shown a tendency to under estimate 4959 coins as 3129.

Let the tendency to overestimate the class of a stimulus be measured by $K' > 0$, the number by which the standard judgment must be multiplied to obtain the mean biased judgment. Let it depend upon $(r_1 - r_0)$, $\cdots$ to $v + 1$ variables thus:

$$K' = K'(r_1 - r_0, \cdots v + 1 \text{ variables})$$
$$\left.\begin{matrix} \psi = 0 \\ \cdots \\ \cdots \end{matrix}\right\} e \text{ equations.}$$

Each set of conditions can thus be defined by a point in $v + 1 - e$ space. If this point be fixed, the conditions of overestimation are constant.

Represent the number of the group determined under standard conditions by r_1. A random judgment under fixed conditions of overestimation will classify the stimulus as r', the mean judgment as r_1' where $K'r_1 = r_1'$. We require to calculate the chance that the stimulus will be identified with r' in a random judgment when the $(v + 1 - e)$ independent variables are fixed and specified by standard judgment as $r_1 \cdots$, etc.

Since experiment has shown a logarithmic function of r' to have significance, consider the following analogue of (7):

$$\sigma_1\sigma = \log_e \frac{r' - r_0'}{r_1 - r_0} \cdot \frac{1}{K'}, \tag{16}$$

where

$$K' = \frac{r_{1+}' - r_0'}{r_{1+} - r_0} \tag{17}$$

and σ_1 is a positive number independent of r'. K' in (17) is seen to be a function of both r_1 and the v's. Since these are fixed σ is a function of r' alone. If the probability of σ can be calculated, the probability of r' follows directly.*

(c) *Probability of* σ

When σ varies from $-\infty$ to ∞, r' varies from r_0' to ∞. Therefore, the postulate of the arithmetic mean is consistent with the definition of σ but not with that of r'. The variables upon which σ depends are doubtless large in number, and this postulate of normality seems likely to be realized. As few as ten independent variables have been shown experimentally to produce a nearly normal distribution. Viewed as a mechanism, each subject certainly brings his private infinity of variables to the aggregate.

The question of independence of these variables is probably incapable of experimental answer. The conditional equations above refer to variables now assumed fixed. Variables within the organism are the dimensions of its history and present condition, many of them independent, some not. Beyond its boundaries are the fluctuating dimensions of environment and those of the stimulus toward which attention is directed. If the judgments are separated by an interval of space or time sufficient in length to assure the absence of direct mutual influence between subjects, the independence of all but a relatively few variables seems likely. The number and variety of all variables is thus also likely to be increased. Even so, apparently independent judgments will be found correlated.[15] Features v of the stimulus and environment as related to the common previous experience of the subjects are probably the source of this correlation. If men respond at all similarly under the same conditions, they will probably respond similarly to a set of biasing external causes. Certain individual tendencies to fancy and bias will tend to be cancelled by opposite tendencies in a large aggregate. The smallness of individual contribution of each variable to deviations in σ is, of course, not assured.

* Since r' can take only discrete values, σ is undefined elsewhere. Experiment is still *represented* by an analytic function of r.

Not to restrict our final result unduly, let us assume that a Gram-Charlier series [16] in the normal law and its derivatives can represent the σ series. Accordingly, the chance that σ take any value is

$$q(\sigma) = \alpha_0\Psi + \alpha_1\Psi' + \cdots, \tag{18}$$

where

$$\alpha_0\Psi = [\sqrt{2\pi}]^{-1} \exp(-\sigma^2/2). \tag{19}$$

If the first term of (18) be predominant, (16) and (19) give for the chance that the judgment r' be less than r:

$$P(r' < r) = [\sqrt{2\pi}\sigma_1]^{-1} \times \int_{r_0'}^{r} \exp\left\{-\frac{1}{2\sigma_1^2}\left[\log_e \frac{r' - r_0'}{(r_1 - r_0)K'}\right]^2\right\} \frac{dr'}{(r' - r_0')}. \tag{20}$$

This is also equal to the expected proportion of the N' judgments less than r, namely n'/N'.

The median and geometric mean about r_0' are both equal to the most probable value of r', namely r_1'. We can thus interpret r_1' of Singer's formula as the solution of

$$\left.\begin{aligned} P(r_1') &= 1/2 \\ r_1' &= r_1K'. \end{aligned}\right\} \tag{21}$$

and K' by

Thus when overestimation obtains, the median judgment shifts from r_1 to $K'r_1$. Similarly,

$$\left.\begin{aligned} P(r_0') &= 0, \\ r_0' &= K'r_0. \end{aligned}\right\} \tag{22}$$

In Singer's definition r_0, r_1, r_{1+} define the unit of the sense scale. A convenient and unambiguous unit for σ is σ_1. For the most part (68.262 percent), the judgments lie between $\sigma = 1$ and $\sigma = -1$. Therefore, when

$$\left.\begin{aligned} \sigma &= 1, \\ P(r_{1+}') &= \frac{1}{2} + 0.34131. \end{aligned}\right\} \tag{23}$$

define r_{1+}' by

Consequently,

$$\sigma_1 = \log_e \frac{r_{1+} - r_0}{r_1 - r_0}. \tag{24}$$

So that finally

$$a = 1/\sigma_1 \tag{25}$$

and a' is given by (16) when $r' = r_1'$, or, by (7). As the conditions leading to overestimation approach standard conditions, the differences between the primed and unprimed quantities become inconsiderable.

(c) *Resumé*

Let us review the method by which (20) was reached. Granted certain explicit a priori principles, matters of fact and of convenience, these are held sufficient to establish a general sensation formula. Every parameter thereof is some stimulus designated by the mean of a type of behavior under constant conditions. The question of appropriate means, *i.e.*, of choice among alternative probability assumptions, is not independent of the form of the sense function. It is a matter of observation that numerical equality judgments can be described by a probability expression involving a function of the same form. Without contradicting its properties, the sense function could be assumed normal in distribution which is not true of the physical variables defining it. The observed distribution of judgments expressed in the r' scale can be deduced if it be assumed that the reciprocals of the sensibilities represented by each class of judgments under conditions of overestimation are normally distributed or nearly so. It leads to the generalized geometric mean and median as the most probable value of an aggregate of such judgments. The generalized geometric mean, the logarithmic standard deviation, and datum of the geometric mean give probability specification to the standard stimulus, limen, physiological zero, and the analogous parameters for overestimation defined by Singer.

Overestimation implies a preferred aggregate N_1 of judgments, here designated by unprimed quantities, with which the aggregate N' is compared. The relations between geometric means r_1, r_1' and their datum stimuli r_0 and r_0' are given by (21) and (22). In general, K' will be a function of $v + 1 - e$ independent variables, r_1 included. It measures the consistency of two aggregates of judgments under different conditions of bias. $K' = 1$ within the sampling error means

probable consistency. The question of the 'true' class of the stimulus need not concern us, for the primed and unprimed terms may be interchanged without affecting the result. Our concern is with consistency of mean judgments and confusion of judgment between neighboring classes.

(*d*) *Comparison of Sensibilities*

Sensibilities of the same kind are those serving identical purposes of the organism. Purpose can be defined as the average result of a class of acts.[17] Here a class of organisms necessarily having nothing more in common than the will to win a prize by the exhibition of sensibility to number has the proximate purpose of classifying a stimulus as one member of a set to which counting can lead, and the immediate object of winning a prize. The mean result r_1 defines these purposes, for on the average $K' = 1$, and behavior in that class can be said to have accomplished its purposes.

The unit of sensibility is that of a standard aggregate exhibiting this kind under constant conditions. The measure of any instance or class of instances is expressed as a multiple σ of the standard $a = 1/\sigma_1$. All relative variations are ascribed to changes in the second, not the first aggregate.

The operations by which sensibility is measured are analogous to those in which lengths are compared. The principal differences are that 'purpose' does not enter the definition of length, nor can standards of sensibility be preserved in exactly the same way. In either case, the constant element is the defined operation and relative variations are not attributed to the standard. Therefore sensibility, like length, is always relative to a standard. That other sensibility we have here designated by the unprimed quantities.

At least two judgments, r_1, r_{1+}, are required to define a standard, for a single judgment does not indicate the order of confusion of classes. As Urban has contended,[18] it is a measure of the standard (or probable) error of that type of classification; in fact, it measures that error in the sense scale. Thus, when aural sensibilities are compared in terms of a unit defined by the average sensibility of the normal ear to various

17

sound energy intensities, the experimenter's sensibility in mechanical classifications of a sound is an inextricable part of the picture. It is the probable error of his mechanical measurements and could serve as a basis for comparing the other sensibilities.

That the nature of the mechanisms involved is immaterial so long as all serve the same end is illustrated by the fact that Gault's Teletactor helps replace the ear with the sense of touch. The sensibility of 'hearing' with the teletactor has been measured in terms of the decibel, a unit of aural sensibility.[19]

A given sensibility can be magnified by suitable choice of physical conditions as illustrated by the telescope and the electrical amplifier. Here estimation and counting represent alternative sets of conditions in which the sensibility to number undergoes indefinitely large magnification.

(*e*) *The Phi-Gamma Hypothesis*

By (11) and (12) when the absolute limen is constant $r_0 = \infty$. If at the same time $(r_1 - r_0)\sigma_1$ remains finite, the distribution is concentrated about r_1'. That is, when

$$\left.\begin{array}{l}(r_1 - r_0) >> 0 \\ (r_1 - r_0)\sigma_1 = s \text{ remains finite}\end{array}\right\} \quad \text{and} \qquad (26)$$

(20) reduces to the Phi-Gamma Hypothesis:

$$P = [\sqrt{2\pi}s]^{-1} \int_{-\infty}^{r} \exp\left\{-\frac{1}{2s^2}\left(r' - r_1'\right)^2\right\} dr', \qquad (27)$$

where r_1' is given by the corresponding limit of (21), and where

$$r_1' - r_1 = (K' - 1)r_1 \qquad (28)$$

has been called the 'constant error' of judgment.

This is an analytic variant of Thurstone's criticism of the Phi-Gamma Hypothesis in which it is shown that (27) is a valid approximation to the facts when discrimination is fine. Since (27) also implies the absolute limen to be constant with the stimulus, it contradicts known fact.

(*f*) *Weber's Law and Other Relations*

The parameters of (20) are independent unless experimental or prior restrictions are imposed on them. By (10) and (14) when r_0 is constant, Weber's Law holds. But returning to (8) as more general and redefining r_2 as in (15), the constancy of r_0 does not imply Weber's Law or any other. We can identify r_{1-}' with the solution of

$$P(r_{1-}') = 1/2 - 0.34131. \tag{29}$$

It is then the stimulus $r' < r_1'$ equally as likely as r_{1+}' to be discriminated from r_1 and the three are in geometric progression with respect to r_0'. The sensibility will not on this definition of r_0 be constant with r_1 except in the special case that $da = 0$, or, where

$$(r_{1+} - r_0) = e^{1/a}(r_1 - r_0), \tag{30}$$

no other restraint being placed upon r_1, r_{1+} and r_0. If interpretations leading to $da = 0$ can be placed upon the r's, a will be a universal constant for that kind of sensibility.

The present experiments lead to interesting consequences if r_0 is fixed at zero value and a allowed to vary with r_1. Solving (6) for r_0

$$r_0 = r_1 \frac{f/r_1 - e^{1/a}}{1 - e^{1/a}}. \tag{31}$$

If

$$\left.\begin{aligned} &r_0 = 0, \\ &r_1 \neq 0, \\ &\text{and } a \text{ finite for all } r_1, \end{aligned}\right\} \tag{32}$$

then

$$\frac{1}{a} = \log_e \frac{r_{1+}}{r_1}. \tag{33}$$

As a special case let

$$\frac{f}{r_1} = \left(\frac{r_1}{b}\right)^{\alpha} \qquad \begin{matrix} \alpha > 0 \\ b > 0; \end{matrix} \tag{34}$$

then

$$\frac{1}{a} = \alpha \log_e \frac{r_1}{b}. \tag{35}$$

If a is to remain positive and finite for $r_1 = 2, 3 \cdots$, etc.,

we must have

$$0 < b \leq 1.$$

Differentiating (34)

$$\frac{df}{dr_1} = (1 + \alpha)\frac{f}{r_1}, \tag{36}$$

which more nearly satisfies (10) when a is small. Whence also

$$r_0 = \frac{\alpha b^{-\alpha} r_1^{1+\alpha}}{(1 + \alpha) r_1^{\alpha} b^{-\alpha} - 1}. \tag{37}$$

Relation between Sensibility and Number of Objects r_1

(*a*) *Experiments II, III, IV*

These experiments employed numbers of objects sufficiently different to make differences in the limen significant.

II: 4959 one cent pieces were exhibited spread level in a rectangular opaque tray 12″ × 18″ × 1″ in the window of a specialty store. Each ten cent purchase of a specialty entitled the customer to one 'guess.' There was no attempt to limit the number of estimates by one person. These were recorded on 3″ × 5″ slips and deposited in a sealed ballot box. The author exercised no control over conditions, merely appropriating the results of the contest.

Of 332 estimates about half were repeats, one person entering 20; in all, 176 could be considered independent. The series shows some preference for numbers which might be classed as common, round, fanciful, lucky, etc. An estimate of 1,000,000 signed 'Amos & Andy,'* was excluded from the calculations.

The data were treated graphically except that the medians and simple geometric means were determined by computation. The latter agreed within the sampling error whether or not the repeats were excluded. Inspection of Fig. 4 will show the effect of excluding arbitrarily 13 repeats by one person. It appears from this that each estimate, whether made by the same or different persons, might be considered as an independent trial though strictly the 332 estimates are not a homogeneous group. The difference between per-

* Currently popular radio comedians.

sons, and the same person on different occasions, is a matter of degree rather than kind. The extreme estimates 56 and 100,000, however hare-brained, have a geometric mean of 2370. The geometric mean of all estimates is 3129 with a standard error of 90, and median, 3065. The count 4959 is some 8 times this error removed from the median. Nevertheless, the form of the law is well preserved especially in the agreement of median and geometric mean, while the arithmetic mean is at 5270. If we assume that when $K' = 1$, $r_{21} = 4959$, we have here an example of underestimation with $K' = .632$, having the same order of error as the geometric mean. The sensibility was lower than for experiment *I*.

III and IV: The cards in Fig. 7 were presented in either order to single subjects. The instructions were: "Without attempting to count them, state the number of circles on this card." The time of exposure was not measured but was of the order of five seconds.

Computation showed agreement between simple geometric and arithmetic means, and median, for both the 11 and the 20 object cases. These were in sufficiently close agreement with 11 and 20 respectively to make a 1 percent over-estimation and 8 percent underestimation inconclusive. The sensibilities exhibited in the four experiments decrease with the number of objects in the group. It was high enough in experiments III and IV to make the Phi-Gamma Hypothesis a fair graduation. These series are shown graphically in Fig. 5. In II, on the other hand, it was so low as to show a marked skewness very plausibly rectified by taking logarithms; Figs. 3, 4.

(*b*) The Function *f*.

As comparability is largely a matter of definition controlled by the sampling error, the fact that the sensibilities in all four experiments are similar makes it permissible to estimate a tentative relation between r_{1+} and r_1. If this relation falls within the sampling error and leads to no absurdity on extrapolation, comparability is fairly well established. Graphical preliminaries lead to (34) as a fair representation, hence to (37) for r_0 on the fourth postulate.

FIG. 3

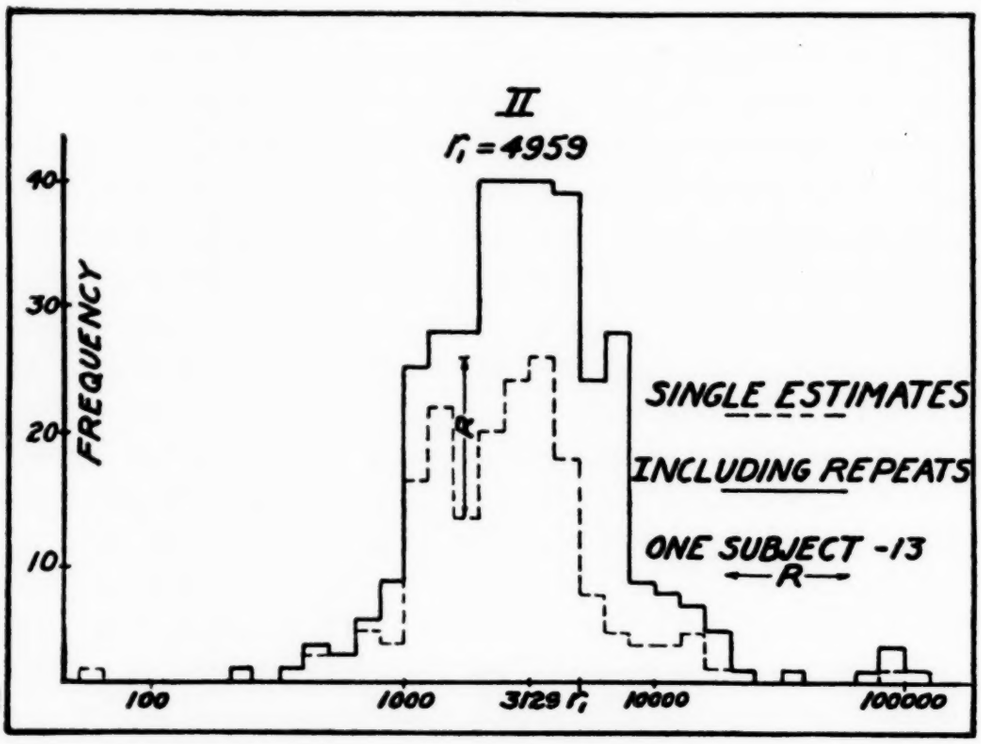

II
$r_1 = 4959$
FREQUENCY
40
30
20
10
SINGLE ESTIMATES
INCLUDING REPEATS
ONE SUBJECT -13
R
100
1000
3129 r_1
10000
100000

FIG. 4

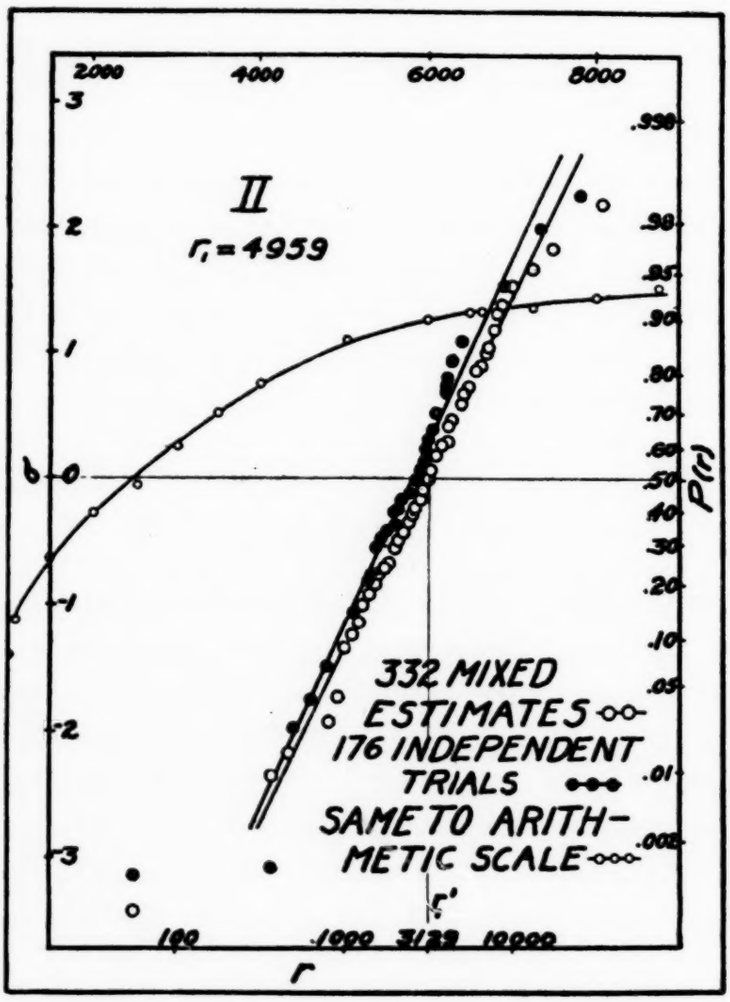

2000
4000
6000
8000
II
$r_1 = 4959$
3
2
1
0
-1
-2
-3
.998
.98
.95
.90
.80
.70
.60
.50
.40
.30
.20
.10
.05
.01
.002
P(r)
332 MIXED ESTIMATES
176 INDEPENDENT TRIALS
SAME TO ARITHMETIC SCALE
r'
100
1000
3129
10000
r

FIG. 5

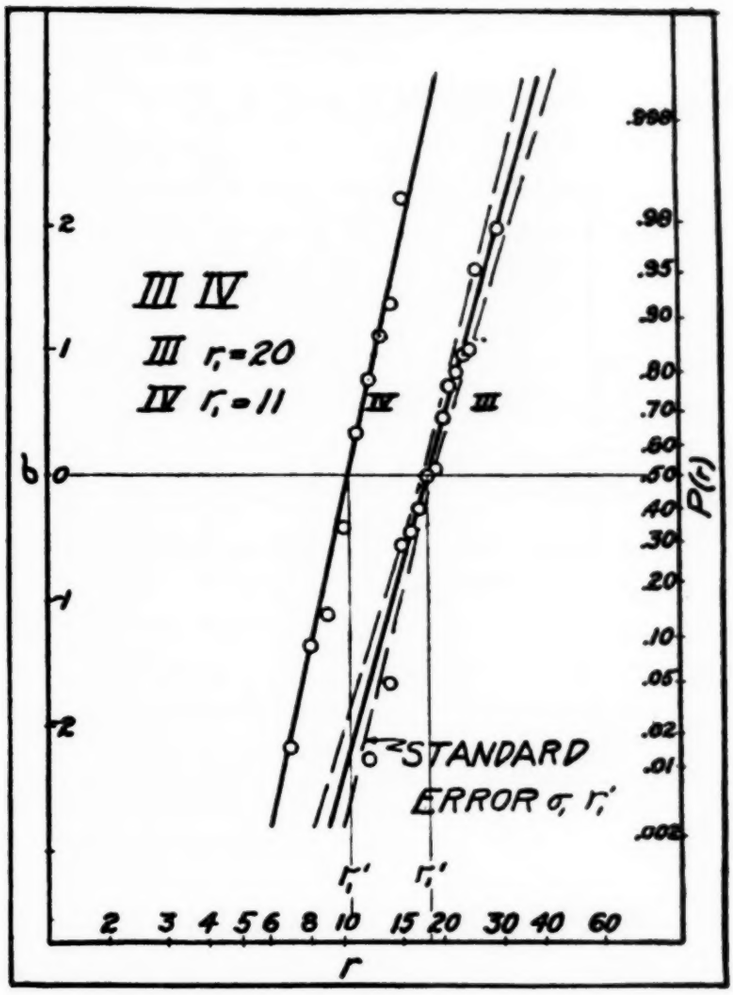

III IV
III r'₁=20
IV r'₁=11
STANDARD ERROR σ r'₁
P(r)
r
2 3 4 5 6 8 10 15 20 30 40 60

FIG. 6

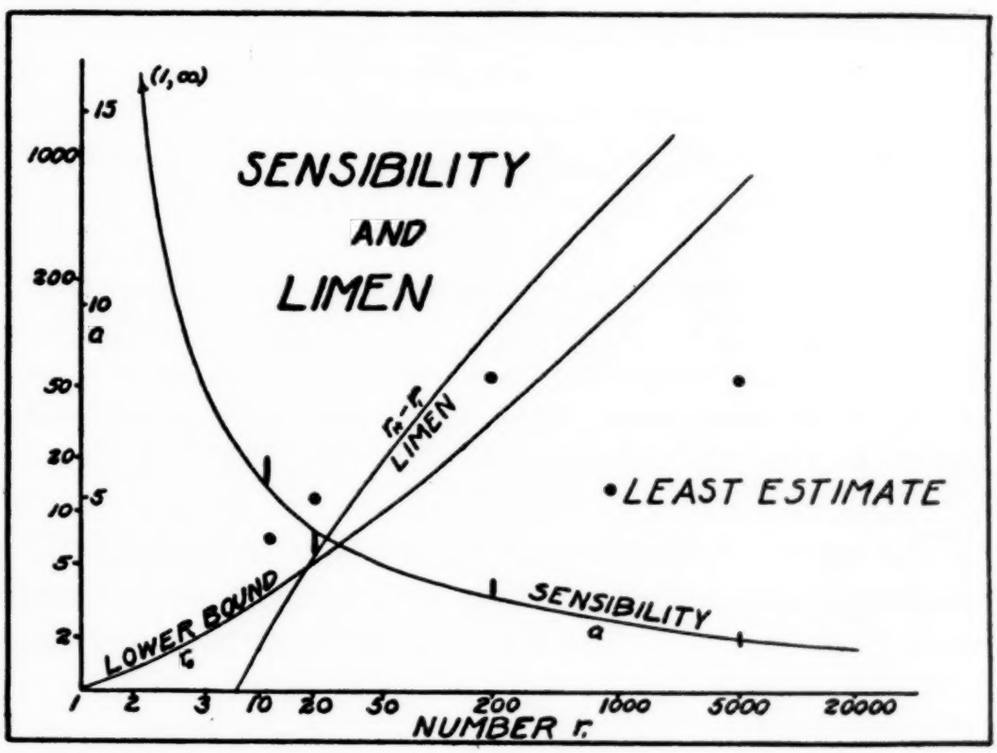

SENSIBILITY
AND
LIMEN
(1, ∞)
LOWER BOUND
r₀
r₀-r
LIMEN
SENSIBILITY
a
LEAST ESTIMATE
NUMBER r

FIG. 7

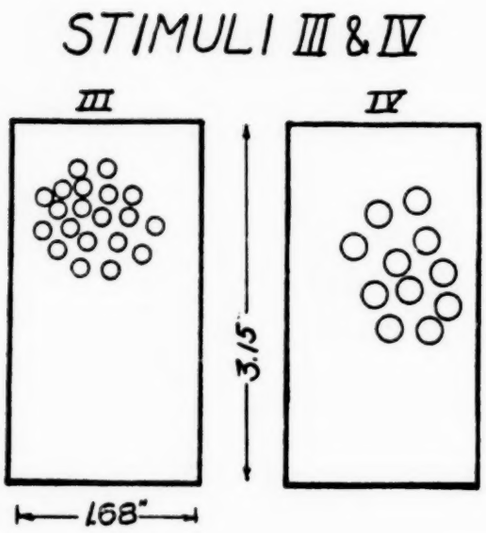


TABLE II

SUMMARY OF DISTRIBUTION PARAMETERS

Stimulus Scale

Exp. No.	r_1	N'	Arith. Mean	$K'r_1$ Geom. Mean	Median	K'	r_1
IV	11 Plane	68	11.1	11.1	10.7*	1.01	0
III	20 "	84	19.2	18.7	17.8*	0.935	0
I	198 Solid	136	204.	191.	191.	0.965	0
I-*A*	198 "	68	195.	183.	191.	0.925	0
I-*B*	198 "	68	213.	200.	190.	1.01	0
II	4959 "	332	5270.	3129.	3065.	0.632	0

ϕ Scale

Exp. No.	$\sigma_1/2.30$	$\log_{10}$ G.M.	$\log_{10}$ Median	$(\sigma_1/2.3\sqrt{N'})$†
IV	0.078	1.0452	1.029*	0.010*
III	0.117	1.2734	1.250*	0.013*
I	0.1569	2.2806	2.2810	0.0135
I-*A*	0.142	2.2622	2.2810	0.0172*
I-*B*	0.166	2.2997	2.2788	0.0201*
II	0.36	3.4954	3.4864	0.0135*

* Starred quantities involve graphical evaluation. Unstarred quantities by computation. Standard error of K' is of same percentage order as for G. M. $= K'r_1$.

† Standard error of σ_1 is .707 times these values.

The lower bounds of the series can be estimated by taking the generalized geometric mean of equi-probable estimates like r_{1-}' and r_{1+}' with the aid of (15) and equations analogous

to (29). It was thus found that r_0' is very probably zero for all four cases. More evidence is that no series conclusively exhibits the characteristic curvature on probability coordinates. If the value of r_0 by (37) be assumed in calculating the geometric mean, agreement with the median is destroyed. Even fairly small assumed values upset this indication that $r_0' = 0$ for all series.

Further strong evidence is given by the excluded estimate 1,000,000. With respect to the simple geometric mean for the group, 3129, its equally probable companion is 9.8. Putting levity aside, our friend would probably regard 10 and 10^6 as equally sensible estimates. But at least the fact of Andrew Brown's willingness to confuse picayunes with millions is too well established not to cast some doubt on his mastery over number. He probably has his correct place in the sensibility scale. This incident makes it likely that as N' increases, all possible members of the set r' will eventually be represented. Therefore, excluding $r_1 = 0$, we must have, whether or not $K' = 1$;

$$r_0 = r_0' = 0. \tag{38}$$

The condition on which postulate 10 was solved for r_ω is thus realized.

But Weber's Law is not verified. If (34) and (38) hold it means that (10) cannot hold. The fourth postulate can in this case be modified by reinterpreting r_2 and abandoning the limiting process. This was done in (15) and (29). Consequently a must be a function of r_1. The sampling error permits the use of (34), hence of (35).

We have now to choose α and b from among experimental possibilities. Employing *I* and *IV* in computation, $b = .58$ and $\alpha = 0.062$. The other five pairs give α a range 0.05 to 0.2 with an arithmetic mean 0.11. The corresponding value of b is 1.06. Since by definition, persons incapable of counting at least one are excluded and since those admitted can probably do so with little or no error, we have $0 < b \leq 1$. The sampling error permits us to take $b = 1$. whence the mean value of α can be calculated and we have:

$$\left.\begin{aligned} \alpha &= 0.083, \\ b &= 1. \end{aligned}\right\} \tag{39}$$

In Fig. 6 are plotted r_0 by (37) and the experimental least estimate; a by (35); and the limen, as functions of r_1 assuming (39). Experimental values of $1/\sigma_1$ with sampling error also appear. The standard sampling error of a is here given by

$$\text{S.E.}(a) = \frac{\text{S.E.}(\sigma_1)}{\sigma_1^2} = 1 \Big/ \alpha(\log_e r_1)\sqrt{N'}. \tag{40}$$

The adopted parameters imply that judgment of unit group is infinitely precise. The chance of error in estimation becomes appreciable for $r_1 = 4$. When $r_1 = 7$, 32 percent of the estimates will be in error by a unit or more. The estimated sensibility curve agrees with the well known fact that counting by more than twos or threes is generally unreliable.*

In conclusion we can say that postulate 4 is here realized provided r_2 be identified by (15) and the limiting process abandoned. The sampling error is too large to justify the rejection of (10) in general, for α is very nearly zero (36). Since the details of the picture presented by the estimated sensibility function are also consistent with common knowledge, the comparability of the four series is in so much well established.

The development of (35) and (39) shows only too clearly that the argument could not have been carried through to that point a priori. The problem is one of restricting assumptions without ignoring relevant facts like the 'Amos and Andy' incident.

Summary

All parameters of Singer's definition of sensation-intensity can find interpretation through the Generalized Law of the Geometric Mean. The postulates of Singer's formula are thus combined with certain probability assumptions. The former are explicit. The latter are open to a familiar class of objections. But all are subject to intelligent modification

* Dantzig [21] tells of the crow whose inability to distinguish between four and five hunters finally led him within reach of the ambushed fifth after having succeeded in estimating correctly 1, 2, 3 and 4.

through experiment. The resulting formula is more general in application than the examples in numerical estimation given here appear to indicate, for mere description is coupled with interesting predictions and new experiments readily suggest themselves. Moreover, apparently independent problems such as the description of overestimation, the connection between Weber's and Fechner's Laws, and the validity of the Phi-Gamma Hypothesis, can be viewed as aspects of the applicability of the law of the geometric mean.

The author takes this opportunity to acknowledge the valuable suggestions of Professor Singer of the University of Pennsylvania, and of Professor M. C. Holmes of West Virginia University.

(Manuscript received November 20, 1932)

Estimates for Experiment I
(In Order Recorded)

150	115	125	350
213	104	101	280
230	200	168	168
228	225	181	147
93	113	275	187
149	115	330	125
275	128	289	150
250	103	199	217
335	299	400	191
257	160	228	181
285	190	250	131
191	225	219	154
248	230	173	175
60	136	206	179
178	147	213	167
221	133	231	129
268	200	143	113
127	107	205	107
178	231	249	135
198	159	196	211
148	256	210	306
147	303	170	325
88	345	224	269
146	320	250	273
132	235	300	288
150	275	150	183
219	190	165	186
248	182	137	234
207	150	177	137
147	175	133	197
348	252	468	156
222	235	643	171
151	263	185	176
118	144 (No. 68)	210	199

Estimates for Experiment II
(Temporal Order Lost)

56	1200	1900	2500*
250*	1212*	1932*	2500*
375	1213*	1932*	2500*
462*	1256	1975	2550
500	1275	1988	2567*
500	1300*	1999	2575*
526	1300	2000	2599
527	1300	2000	2600
650	1313*	2000	2600
687*	1313*	2000	2618*
702	1313*	2000	2675
750	1313	2000*	2675
775	1369*	2013	2700*
904*	1375	2024*	2700
932*	1383	2090*	2777
963	1400	2097*	2777
983*	1436	2100	2798
993*	1498	2111*	2800*
997*	1499	2113*	2834*
998	1500	2115*	2845
999	1500	2121	2863*
1000	1500	2137	2875
1001	1500	2150	2900
1001	1500	2165*	2940
1001*	1500	2173	2949
1005	1532*	2175	2962*
1005	1533	2217*	2972*
1009*	1549	2225	2981*
1017*	1550	2262*	2998
1024	1525	2305*	2999*
1027*	1526	2323*	2999
1050	1572	2327*	3000
1097	1600	2333	3000
1099	1600	2349*	3000
1099	1625	2375	3000
1111	1637	2404*	3000*
1111*	1675*	2482*	3015
1118	1700	2492*	3019*
1137*	1710*	2499	3050
1149	1739	2500	3060*
1156	1763	2500	3072
1200	1800	2500	3072*

3158*	4000	5169*	7843*
3163*	4000*	5238*	7850*
3166*	4000*	5272*	7855*
3187	4027	5331*	7878*
3190	4032	5351	7888*
3200	4052*	5401*	7898*
3215	4060*	5467*	8000
3217*	4145	5492	8100*
3243	4187	5500	8125*
3247*	4200	5555	8191*
3300	4200	5601*	8500
3333*	4201*	5695*	8750*
3333*	4250	5700*	8763
3335	4272*	5711	9894*
3400	4300	5865*	9999*
3450	4300	6000	9999*
3456*	4320*	6001*	10002
3456	4327	6093*	10101
3475	4428	6219*	10630*
3500	4475	6233*	12000*
3500	4500*	6400	12500
3500	4500*	6478*	12750
3500	4500*	6500	13000
3500	4500*	6666*	13500*
3540*	4610	6780*	13737
3550	4750	6875*	14560*
3599*	4753	6917*	15000
3601*	4763*	6932*	16280*
3650*	4800	6942*	16425*
3660*	4800*	7200*	17800
3773*	4938*	7200*	18200*
3750	4980*	7222*	25000
3750	4996*	7253	35030*
3750	4999*	7341*	70013*
3792	5000*	7434*	80013*
3860	5000*	7437*	88013
3890*	5000*	7500	90013*
3896	5000*	7500*	100099
3950	5027*	7501*	
3951	5031*	7575*	
3955	5050*	7600*	
4000	5150*	7658*	

(1,000,000 excluded—see text)
Unstarred numbers are independent estimates.

Estimates for Experiment III

Estimate	Frequency
12	1
13	0
14	3
15	20
16	3
17	6
18	9
19	2
20	13
21	7
22	3
23	3
24	1
25	9
26	0
27	1
28	1
29	0
30	2
	84

Estimates for Experiment IV

Estimate	Frequency
6	0
7	1
8	5
9	3
10	14
11	20
12	10
13	6
14	4
15	4
16	0
17	0
18	1
	68

References

1. Note in: Whitaker & Robinson, *Calculus of Observation*, p. 217, Blackie, London, 1924.
2. Galton, F., *Proc. R. Soc.*, 1879, **29**.
   McAlister, D., Law of the geometric mean, *Proc. R. Soc.*, 1879, **29**, p. 367.
3. Fechner, G. T., *Masskollektivlehr*, Leipzig, 1897.
4. Kapteyn, J. C., Skew frequency curves; First Paper, Groningen, 1903, Stechert's, N. Y.
   Kapteyn, J. C. and Van Uven, M. J., Second Paper, Astronomical Laboratory, Groningen, 1916.
   Van Uven, M. J., Skew frequency curves, Sec. of Sci., Proc. Kon. Akad., Amsterdam, XIX (No. 4), 1916.
5. Pearson, K., Das Fehlergesetz und seiner Verallgemeinerungen durch Fechner und Pearson, *Biometrika*, 1905–06, **4**, p. 169.
   Pearson, K., A rejoinder to Prof. Kapteyn, *Biometrika*, 1906–07, **5**, p. 168.
6. Hertzsprung, A., *Z. f. Wiss. Phot.*, **3**, p. 468, 1905.
7. Shewhart, W. A., *Bell System Tech. Jour.*, Jan. 1924, p. 59.
8. Thurstone, L. L., The phi-gamma hypothesis, Jour. Exper. Psychol., Vol. XI, Aug. 1928, pp. 293–305.
9. Fritz, F. F., Experimental evidence in support of Prof. Thurstone's criticism of the phi-gamma hypothesis, *Jour. Gen. Psych.*, Dec. 1930, p. 346.
10. Singer, E. A., *Mind as Behavior*, Ch. VI; Adams, Columbus, O., 1924. Also; Jour. Phil., June 1917.
11. *Jour. Am. Stat. Assoc.*, Dec. 1929, p. 349.
12. Hecht, S., *J. O. S. A. & R. S. I.*, **18**, March 1929, pp. 274–278.
13. Wever, E. G. and Zener, K. E., Method of absolute judgment in psychophysics, *Psych. Rev.*, 1928, **35**, p. 466.
14. Ponzo, M., Urteilstaüschung über Mengen, *Arch. f. d. ges. Psych.*, 1928, **65**, p. 129.
15. Pearson, K., On the mathematical theory of the errors of judgment with special reference to the personal equation, *Phil. Trans. R. Soc.*, London, Vol. 198-A, pp. 235–299.
16. Fisher, A., *Math. Theory of Prob.*, Ch. 13, 14.
17. Ref. 10, Ch. IV.
18. Urban, F. M., *Psych. Rev.*, **17**, 1910, p. 229.
19. Doan, G. E., *Jour. Frank. Inst.*, Sept. 1933, p. 351.
20. Dantzig, T., *Number*, p. 3, MacMillan, 1930.

VISUAL PERCEPTION OF HORIZONTAL MOVEMENT [1]

BY RALPH O. VAN WATERS
Psychological Laboratory, University of Chicago

INTRODUCTION

When two lights are presented in a dark room and one light is made to move while the other is stationary, an observer is subject to considerable error in his judgments. The frequency and degree of error vary with circumstances. This phenomenon was first studied by Carr and Hardy [2] and it was found by them that a variety of circumstances including absolute and relative intensity, size and possibly fixation influenced the amount of error. With an improved form of apparatus the phenomenon was studied by Mayberry whose results are unpublished. Thelin [3] found that the degree of error was dependent upon the intensity and form of the two lights.

Our apparatus was the same as that employed by Thelin with but one exception: the two light filled boxes were so arranged that the two lights were placed in the same plane and equally distant from the observer. The apparatus has been described with such detail that duplication is unnecessary.

The following constant conditions obtained throughout the experiment. The subject was seated so that the lights were approximately at the height of his eyes. The subject observed the lights at a distance of approximately 5 feet. During a single exposure one light was fixated continuously. Both lights were equal in intensity. Both lights were circular in shape. The moving light always moved away from the stationary light, and never toward it. The amount of movement was always 5 inches. At the beginning of movement the two lights were separated by a distance of $1\frac{3}{4}$ inches measuring from center to center of the lights. At the end of the movement a distance of $6\frac{3}{4}$ inches separated the lights.

At regular intervals throughout the experiment both lights were made to move in order to induce in the subject an attitude favorable to the illusory effect. In this case

[1] The author is indebted to Harvey A. Carr for valuable suggestions in organizing the research and for patience in correcting this paper.

[2] Carr, H. A. & Hardy, M. C., Some factors in the perception of relative motion; a preliminary experiment, *P. R.*, 1920, **27**, 24–37.

[3] Thelin, E., The perception of relative visual motion, JOUR. EXPER. PSYCHOL., 1927, **10**, 321–348.

the lights moved apart at the same rate of speed and each light moved for a distance of 5 inches. The exposure time was thus the same as in the standard conditions. These supplementary data were separately tabulated and studied. The data are not included because they differ in no respect from those given in this paper.

The phenomenon was studied under the following variable conditions:

1. Speed: Three rates of movement were employed; namely, speeds of $\frac{1}{4}$ inch per second, $\frac{1}{2}$ inch per second, and 1 inch per second.

2. Size: Three sizes were used. These were circles of a diameter of 1 inch, of $\frac{1}{2}$ inch, and of $\frac{1}{4}$ inch. Combinations of these sizes as they appeared were: (I) the 1 inch circle with the $\frac{1}{4}$ inch circle, (II) the 1 inch circle with the $\frac{1}{2}$ inch circle, and (III) the $\frac{1}{2}$ inch circle with the $\frac{1}{4}$ inch circle. All three speeds were used in conjunction with the three size combinations.

3. Fixation: (*a*) In 50 percent of the exposures the small light was fixated, in 50 percent of the exposures the large light was fixated; (*b*) In 50 percent of the cases the moving light was fixated, in 50 percent of the cases the stationary light was fixated; (*c*) In 50 percent of the exposures the right light was fixated, in 50 percent of the exposures the left light was fixated.

4. Direction of movement and Position: (*a*) In 50 percent of the cases the moving light travelled to the observer's right, in 50 percent of the cases the moving light travelled to the observer's left; (*b*) In 50 percent of the exposures the small light was on the right, in 50 percent of the exposures the small light was on the left.

5. Incidentally the effects of practice were noted throughout the experiment.

The conditions for size, fixation, movement, and position, with speed constant, give the following eight combinations:

1. Small light on right; small fixated; large only moves.
2. Large light on right; large fixated; small only moves.
3. Small light on right; small fixated; small only moves.
4. Large light on right; large fixated; large only moves.
5. Small light on right; small fixated; large only moves.
6. Large light on right; large fixated; small only moves.
7. Small light on right; small fixated; small only moves.
8. Large light on right; large fixated; large only moves.

Since the number of size variables is three, and the number of speed variables is also three, each of the above combinations can be subdivided into the following nine combinations:

1. Smallest light with largest light and slowest speed.
2. Smallest light with medium light and slowest speed.
3. Medium light with largest light and slowest speed.
4. Smallest light with largest light and medium speed.
5. Smallest light with medium light and medium speed.
6. Medium light with largest light and medium speed.
7. Smallest light with largest light and fastest speed.
8. Smallest light with medium light and fastest speed.
9. Medium light with largest light and fastest speed.

This makes a total of 72 possible conditions of exposure. Each of these conditions was presented 10 times to each of the ten subjects, but we must bear in mind that, disregarding direction of fixation, we have received 20 reactions from each of the subjects on all conditions. Hence, our data cover a grand total of 7,200 recorded reactions.

The following instructions were given our subjects: "You will be seated in a dark room to observe lights. There will be two of them. One will always be larger than the other. Only one light will appear at first and you are to fixate or observe this one constantly. Remember, no matter what else happens do not take your eyes off this light. You are to report any movement which you see. The lights will never move together. With more and more time they will be farther and farther apart. Now I want you to report EXACTLY WHAT YOU SEE. If both lights appear to be moving at the same speed, report that. If both appear to be moving apart with one moving faster, report that. If one light appears to be doing all of the moving, report that. If both appear to be moving in the same direction with one moving faster than the other so that they are getting farther and farther apart, report that. ONLY REPORT EXACTLY WHAT YOU SEE."

Prior to the beginning of the light exposures the subject was required to sit in the dark room until visual adaptation had occurred. The ability to see an extremely dim light served as the criterion that adaptation was complete enough that our experiment might proceed. The time required for adaptation varied from subject to subject and from time to time, but in most cases from five to ten minutes was the maximum time required.

One of the two lights was then presented to the subject and it was understood to be the light for constant fixation. For two seconds this light remained within his field of vision and at the end of that period the second light appeared. Movement began simultaneously with the appearance of the second light. At the end of movement both lights were automatically extinguished and the judgment was then made and recorded.

Ten subjects were used in the experiment, 6 men and 4 women. Three of these were students of psychology and were somewhat acquainted with general laboratory technique. None was aware of the purpose of the experiment, and at no time were the subjects permitted to check their judgments against objective conditions.

SECTION I

Effect of Various Variable Factors on Accuracy

To what extent is the accuracy with which the subjects see this situation dependent upon each of the variable conditions? Of all possible judgments given by the individual subjects it is possible to classify these reports with respect to their accuracy in terms of the following six categories:

1. A judgment which reports precisely the actual physical conditions and is therefore *completely accurate.*
2. A judgment which reports the lights as moving in opposite directions at the same speed, and is therefore *partially accurate.*
3. A judgment which reports the lights as moving in opposite directions with the larger light moving faster and is therefore *partially accurate.*
4. A judgment which reports the lights as moving in opposite

directions with the smaller light moving faster and is therefore *partially accurate.*

5. A judgment which reports the lights as moving in the same direction with one light moving faster and is therefore *partially accurate.*
6. A judgment which reports precisely the opposite of the actual physical conditions and is therefore *completely inaccurate.*

For ease and convenience of tabular presentation and treatment these judgments have been further reduced to the categories of CA, PA, and CI (complete accuracy, partial accuracy, and complete inaccuracy). All accuracy tables give the frequency of the three degrees of accuracy in percentage terms.

It is obvious that the category of partial accuracy may be further analyzed in respect to degree of accuracy. Such analysis has always been made and when significant results have been revealed, these data have been presented in supplementary tables.

As a measure of reliability we have used Yule's [4] formula for determining standard error of the differences of observed proportions, but this has been submitted only where the data are made more intelligible by it and with most cases where the results have been concerned with group averages. In a good many instances the reliability measure is not included because the subjects have shown a consistent tendency which is apparent upon inspection. A sufficient degree of reliability is assumed when Diff./S.E. (diff.) has been found to be 3 or more.

Speed

Is the degree of accuracy a function of the speed of movement? In the analysis here our sole variable is speed; all other factors such as size, position, fixation, etc., are held as constants by virtue of balancing all conditions, *e.g.*, in half of the instances the moving light is the smaller light, in the other half of the cases the moving light is the larger light. These

[4] Yule, G. U., *An Introduction to the Theory of Statistics.* London: Griffin, 1927, 268-271.

figures on speed, and indeed all figures presented in the paper are expressed in percentage terms which have been calculated from total frequencies in any category. For the sake of brevity we have indicated in the tabular arrangements "*A*" as the slowest speed, "*B*" as the medium speed, and "*C*" as the fastest speed. The factual data have been treated under the three comparisons: (1) *A* and *C*, (2) *B* and *C*, and (3) *A* and *B*.

TABLE I*a*

SPEED EFFECTS ON ACCURACY

Comparison of Slowest Speed with Fastest Speed

	A (Slowest Speed)	*C* (Fastest Speed)	Diff.	S.E. Diff.	$\frac{D}{\text{S.E. Diff.}}$
C.A.....	23.84	37.17	13.33	1.31	10.2
P.A.....	68.20	61.96	6.24	1.37	4.6
C.I.....	7.96	.87	7.09	.58	12.2

Group results. 4,800 reactions.

It is apparent upon inspection of Table I*a* that complete accuracy is greater, partial accuracy less, and complete inaccuracy less for the faster speed. These differences are reliable. The reliability of the differences of these group averages is further attested by the fact that there is but one exception in the individual records for all possible comparisons, this being a subject whose records do not agree with group results in the P.A. differences.

TABLE I*b*

SPEED EFFECTS ON ACCURACY

Comparison of Medium Speed with Fastest Speed

	B (Medium Speed)	*C* (Fastest Speed)	Diff.	S.E. Diff.	$\frac{D}{\text{S.E. Diff.}}$
C.A.....	26.87	37.17	10.31	1.34	7.7
P.A.....	69.42	61.96	7.46	1.37	5.4
C.I.....	3.71	.87	2.84	.43	6.6

Group results. 4,800 reactions.

When the medium and fastest speeds are compared, we find that the higher speed also shows higher C.A., lower P.A., and lower C.I. Differences are again reliable, and like the comparisons of speeds *A* and *C*, we find but one exception in

the individual records, this being in a comparison of the P.A. category.

The results of the comparison of the slowest speed with the medium speed are shown in Table I*c*. In the categories C.A. and P.A. there are no significant differences for group results. Our differences do not show the reliability found in the previous comparisons. But complete inaccuracy is significantly lower for the faster speed. Individual records

TABLE I*c*

SPEED EFFECTS ON ACCURACY

Comparison of Slowest Speed with Medium Speed

	A (Slowest Speed)	*B* (Medium Speed)	Diff.	S.E. Diff.	$\frac{D}{\text{S.E. Diff.}}$
C.A.....	23.84	26.87	3.03	1.25	2.4
P.A.....	68.20	69.42	1.22	1.34	.9
C.I.....	7.96	3.71	4.25	.64	6.6

Group results. 4,800 reactions.

which are not contained in Table I*c* show that 7 of the 10 subjects have higher complete accuracy for the medium speed, 5 of the 10 subjects have greater partial accuracy for the medium speed, and all 10 subjects are consistent in showing less complete inaccuracy for the medium speed.

Analysis of Partial Accuracy.

Obviously when a subject reports two lights to be in motion when only one is actually in motion, he is more nearly correct in his judgment if he reports the stationary light moving slower than if he should report the lights moving apart at the same rate of speed, or the stationary light actually moving faster than the moving light. Our partially accurate judgments, then, may be placed in the following relative order of accuracy:

1. Stationary light reported moving slower than the moving light. (Greatest accuracy.)
2. Stationary light reported moving at the same speed as the moving light. (Medium accuracy.)
3. Stationary light reported moving faster than the moving light. (Least accuracy.)

Table II presents judgments arranged in this manner.

It can be seen from Table II that accuracy is increased, inaccuracy decreased with faster speeds. All three accuracy categories show this tendency. This is true for the comparisons of *A* with *C*, *B* with *C*, and *A* with *B*. The reliability

TABLE II

SPEED EFFECTS ON ACCURACY

Analysis of Partial Accuracy

	Speed *A*	Speed *B*	Speed *C*
Greatest Accuracy	38.79	51.10	69.76
Medium Accuracy	39.47	33.66	22.46
Least Accuracy	21.74	15.24	7.78

Group results. Total partially accurate judgments—4,790.

of the group comparisons is attested by the consistency of the individual records which are not given. There are but 4 exceptions in the 30 comparisons in the first category, greatest accuracy. Of the 30 comparisons for medium accuracy, 27 support the group results, and a like number of 27 support the third category, least accuracy.

Summing up the effects of the variable factor, speed, on the accuracy of perception, we have shown that as speed is increased accuracy increases also. Increasing the speed increases complete accuracy and decreases partial accuracy and complete inaccuracy. Individual records consistently substantiate these group findings.

Size

The effects of size on accuracy of judgment can best be stated in the form of answers to the following questions: (1) How is accuracy affected by a great difference in size between the two lights; by a medium difference in size; by a small difference in size? (2) Will a larger, medium or smaller combined size or total retinal stimulation differentially affect accuracy?

Difference in Size.

There are three size combinations in which a large and a small light appear together. These shall be designated as I, II, and III:

Pair I—$\frac{1}{4}$ inch circle with the 1 inch circle, or the greatest difference in size. ($\frac{3}{4}$ inch.)

Pair II—$\frac{1}{2}$ inch circle with the 1 inch circle, or the medium difference in size. ($\frac{1}{2}$ inch.)

Pair III—$\frac{1}{4}$ inch circle with the $\frac{1}{2}$ inch circle, or the least difference in size. ($\frac{1}{4}$ inch.)

The answer to the first question is to be had by examination of Table III.

TABLE III

SIZE EFFECTS ON ACCURACY

Difference in Size

	Pair I	Pair II	Pair III
C.A.	29.17	29.59	29.12
P.A.	65.83	66.24	67.51
C.I.	5.00	4.17	3.37

Group results. 7,200 reactions.

There are no significant differences between any of the categories, and the S.E. difference calculations are omitted. Examination of individual records both for the above categories and partial accuracy did not reveal any significant influences of difference in size upon accuracy.

Total Retinal Stimulation.

It is apparent that pair II offers the greatest retinal stimulation ($1\frac{1}{2}$ inches), pair I the medium area of stimulation ($1\frac{1}{4}$ inches), and pair III the least amount of stimulation ($\frac{3}{4}$ inches). Hence, it is possible to answer the second question also by inspection of Table III. No reliable group differences are observable and individual records likewise are insignificant.

We must conclude that the extent of total retinal stimulation, within the limits employed, has no effect upon illusion or degree of accuracy.

Fixation

Moving Light vs. Stationary Light.

The stationary light and the moving light were each fixated continuously in 50 percent of the exposures. What effect on perceptual accuracy can be noted from fixation of a moving or a stationary light?

Group results found in Table IV*b* are pronounced in indicating that complete accuracy is greater, partial accuracy less, and complete inaccuracy less when the moving light is fixated. Hence, fixating a moving light is conducive to maximum accuracy.

Individual results have been presented to point out any variations that may exist. Eight of the ten subjects are consistent in regarding the fixated moving light with greater accuracy. With subjects AU and WA, however, there is apparently no effect on accuracy of this factor. It would be an error to regard this evidence as conclusive until the results of partial accuracy have been analyzed. Both subjects have contributed a large percentage of their judgments to this category and if any effects are apparent at all they should be found in an analysis of partial accuracy.

TABLE IV*a*

FIXATION EFFECTS ON ACCURACY

Moving Light Fixated vs. Stationary Light Fixated

	Moving Light Fixated			Stationary Light Fixated		
	C.A.[1]	P.A.[1]	C.I.[1]	C.A.[2]	P.A.[2]	C.I.[2]
ZE	18.60	77.50	3.90	16.40	80.00	3.60
GR	11.40	88.60	0	3.32	96.68	0
CU	62.50	35.00	2.50	40.55	53.35	6.10
HA	86.68	10.00	3.32	53.90	22.50	23.60
AU	6.68	93.32	0	6.95	92.77	.28
OK	74.18	19.15	6.67	67.22	25.28	7.50
WA	6.67	93.05	.28	6.67	91.65	1.68
TA	29.17	70.83	0	5.82	91.95	2.23
WH	58.60	40.00	1.40	14.72	69.73	15.55
HO	12.77	86.95	.28	3.05	92.22	4.73

Individual results. 7,200 reactions.

TABLE IV*b*

FIXATION EFFECTS ON ACCURACY

Moving Light Fixated vs. Stationary Light Fixated

	Moving Fixated	Stationary Fixated	Diff.	S.E. Diff.	$\frac{D}{\text{S.E. Diff.}}$
C.A.	36.72	21.85	14.87	1.05	14.2
P.A.	61.45	71.62	10.17	1.10	9.2
C.I.	1.83	6.53	4.70	.47	10.0

Group results. 7,200 reactions.

Analysis of Partial Accuracy.

TABLE V

FIXATION EFFECTS ON ACCURACY

Fixation of Moving Light vs. Fixation of Stationary Light

Analysis of Partial Accuracy

	Moving Fixated	Stationary Fixated
G.A.	65.87	41.34
M.A.	27.62	36.08
L.A.	6.51	22.58

Group results. Total partially accurate judgments, 4,790.

Group results in Table V show that accuracy of perception is increased when the moving light is fixated.

Table VI presents individual records of subjects AU and WA. No other individual records are presented because their results are consistent with group findings which appear in Table V and AU and WA were individual exceptions above in Table IV*a*.

TABLE VI

FIXATION EFFECTS ON ACCURACY

Fixation of Moving Light vs. Fixation of Stationary Light

	Moving Fixated	Stationary Fixated	Diff.	S.E. Diff.	$\frac{D}{\text{S.E. Diff.}}$
Subj. AU					
G.A.	66.67	65.16	1.51	3.66	.4
M.A.	21.13	16.52	4.61	3.01	1.5
L.A.	12.20	18.32	6.12	2.77	2.2
	$N = 336$	$N = 333$			
Subj. WA					
G.A.	74.93	59.70	15.23	3.60	4.2
M.A.	16.71	20.61	3.90	3.02	1.3
L.A.	8.36	19.69	11.33	2.66	4.3
	N = 335	$N = 330$			

Differences are not reliable for subject AU, and we must conclude that, with this individual, accuracy is unaffected by fixation of a moving *vs.* a stationary light. But with WA we find that there is considerable effect—and that in the direction of the other subjects; namely, that a moving light fixated is conducive to maximum accuracy and that errors of perception are most apparent when the stationary light is fixated.

Small Light vs. Large Light.

As another aspect of fixation we must consider fixation of a small light as opposed to fixation of a large light. Will accuracy be greater when the small light is fixated? Table VII presents group results.

TABLE VII

FIXATION EFFECTS ON ACCURACY

Fixation of Small Light vs. Fixation of Large Light

	Small Light Fixated	Large Light Fixated
C.A.	29.11	29.47
P.A.	65.39	67.67
C.I.	5.50	2.86

Group results. 7,200 reactions.

There is no significant difference between fixation of a small light and fixation of a large light in the categories of 'complete accuracy' and 'partial accuracy.' But with 'complete inaccuracy' the difference is significant. This tendency, that the small fixated light is conducive to more complete inaccuracy, finds no individual exceptions among the ten subjects.

We present no results here for partial accuracy analysis. Suffice it to say that the category 'least accuracy' is greater for the small fixated light than for the large fixated light, and that this difference remains as the only significant difference between the categories of partial accuracy.

Right Light vs. Left Light.

Group results on fixation of a right light as opposed to fixation of a left light are presented in Table VIII.

TABLE VIII

FIXATION EFFECTS ON ACCURACY

Fixation of Right Light vs. Fixation of Left Light

	Right Light Fixated	Left Light Fixated
C.A.	28.50	30.05
P.A.	67.31	65.78
C.I.	4.19	4.17

Group results. 7,200 reactions.

Fixation of the right light or left light has no effect upon accuracy of perception. Not only are the differences in Table VIII unreliable, but 8 of the 10 subjects show no significant differences in the categories. With subjects WA and TA, however, there is some effect which is apparent upon inspection of the regular accuracy categories and an analysis of partial accuracy. WA displays greater accuracy in fixating the right light, TA manifests greater accuracy in fixating the left light. These results are not presented here, but the differences are just barely reliable.

It may be said that direction of fixation has no effect on accuracy for the majority of subjects.

Position and Direction of Movement

Position.

Will accuracy be greater by having the small light on the right or left of the observer? Group results are presented in Table IX because there were no individual variations significant enough to record.

TABLE IX

POSITION EFFECTS ON ACCURACY

	Small Light on Right	Small Light on Left
C.A.	29.86	28.72
P.A.	66.14	66.92
C.I.	4.00	4.36

Group results. 7,200 reactions.

No significant differences occur in group results for position. No individuals gave differences which were reliable.

Nine of the ten subjects gave nothing significant for position through analysis of partial accuracy. One individual, AU, tended to regard the situation with the small light on the left more accurately than the condition in which the small light was on the right.

Direction of Movement.

Is accuracy greater when the moving light is traveling to the right or left? This tendency to judge one or the other directions of movement more or less accurately has been found

to be largely a matter of individual difference. Diversity of response from the subjects has made group results worthless so that no group differences have been found to be reliable. For this reason in Table X we present individual records so that the effect of the factor direction of movement on accuracy may be noted.

TABLE X

DIRECTION OF MOVEMENT EFFECTS ON ACCURACY

Movement to Right vs. Movement to Left

	Movement to Left			Movement to Right		
	C.A.[1]	P.A.[1]	C.I.[1]	C.A.[2]	P.A.[2]	C.I.[2]
ZE	23.89	75.55	.56	11.11	81.95	6.94
GR	8.06	91.94	0	6.67	93.33	0
CU	50.00	44.72	5.28	53.06	43.61	3.33
HA	64.34	16.95	18.61	76.11	15.56	8.33
AU	3.33	96.67	0	10.28	89.44	.28
OK	75.28	19.44	5.28	66.11	25.00	8.89
WA	7.50	91.39	1.11	5.84	93.33	.83
TA	18.62	79.44	1.94	16.39	83.33	.28
WH	39.17	54.16	6.67	34.17	55.55	10.28
HO	11.94	86.67	1.39	3.89	92.50	3.61

Individual results. 7,200 reactions.

Accuracy is unaffected for 5 of the 10 subjects by direction of movement. Of the other 5, ZE, OK and HO display greater accuracy when the movement is to the left, HA and AU regard movement to the right with greater accuracy. The reliability of these differences is not recorded here because partial analysis has yielded us with evidence which proves the data to be significant.

Analysis of Partial Accuracy.

Table XI presents the data of the five affected subjects for partial accuracy. The other five subjects have not been included because, as above, they have been unaffected by direction of movement.

The data of partial accuracy make the results of direction of movement conclusive; that, five of the ten subjects are unaffected, three perceive movement to the left more accurately, and two perceive movement to the right more accurately.

TABLE XI

DIRECTION OF MOVEMENT EFFECTS ON ACCURACY

Movement to Right vs. Movement to Left

Analysis of Partial Accuracy

	Movement to Left			Movement to Right		
	G.A.[1]	M.A.[1]	L.A.[1]	G.A.[2]	M.A.[2]	L.A.[2]
	N =			N =		
ZE...	(272) 47.79	47.79	4.42	(295) 34.58	52.20	13.22
HA...	(61) 60.66	0	39.34	(56) 71.43	0	28.57
AU...	(348) 54.03	23.56	22.41	(322) 78.88	13.66	7.46
OK...	(70) 12.86	87.14	0	(90) 11.11	84.45	4.44
HO...	(312) 47.76	40.70	11.54	(333) 40.54	44.14	15.32

Individual results. Total partially accurate judgments, 2,159.

Practice

Since our subjects were allowed no opportunity to check their judgments against actual physical conditions, one would expect little or no improvement in accuracy through practice. To check, however, on the possibility that learning might have taken place the data have been arranged to show any changes in accuracy at the end of the 2d fifth, the 3d fifth, and the 5th fifth of the experiment. The group results are shown in Table XII*a*.

TABLE XII*a*

PRACTICE EFFECTS ON ACCURACY

End of 2d Fifth	End of 3d Fifth	End of 5th Fifth
C.A.[1] 31.49	C.A.[2] 27.64	C.A.[3] 27.57
P.A.[1] 63.86	P.A.[2] 69.03	P.A.[3] 68.37
C.I.[1] 4.65	C.I.[2] 3.33	C.I.[3] 4.06

Group results. 7,200 reactions.

TABLE XII*b*

PRACTICE EFFECTS ON ACCURACY

Reliability Table

	C.A.[1]–C.A.[3]	P.A.[1]–P.A.[3]	C.I.[1]–C.I.[3]
Difference	3.92	−4.51	.59
S.E. Difference	1.20	1.25	.54
$\frac{D}{\text{S.E. Diff.}}$	3.3	3.6	1.1

Group differences are very small throughout various stages of the experiment, but from Table XII*b* we can see that what differences exist show that there is a slight dropping off of accuracy and an increase of the illusion with practice for the group as a whole. These differences are barely reliable.

Individual records showed that but one subject, HA, tended to increase accuracy with practice, and the differences in this instance are just significant. On the other hand, there were three subjects who displayed rather marked tendencies to lower their accuracies; these are GR, AU, and HO. We may conclude that there is no evidence to show that practice tends to decrease the illusion; if anything, there is a slight tendency for individuals to make rather stereotyped judgments and persist in certain errors throughout practice.

Summary: Factors Affecting Accuracy.

Factors tending to increase accuracy are:

(1) Increased speed.
(2) Fixation of a moving light.

Factors tending to decrease accuracy are:

(1) Decreased speed.
(2) Fixation of a stationary light.
(3) Fixation of a smaller light.

Factors ineffective in influencing accuracy are:

(1) Increase or decrease in the difference in size between the larger and the smaller light.
(2) Increase or decrease in the total amount of retinal stimulation.
(3) Direction of fixation. (Group results.)
(4) Direction of movement. (Group results.)
(5) Position of a certain light either to the right or to the left of the observer.
(6) Practice. (Excluding individual variation.)

Section II

The Relative Mobility of the Two Lights

The term mobility refers to the susceptibility of the lights to the illusory effect, the tendency to see each light as moving

when it is stationary. Each light actually moved in 50 percent of the cases. Our data show that both lights were perceived in motion in more than 50 percent of the cases, and hence both lights were mobile.

When the two lights are different in some respect, such as size, fixation, or position, are they equally susceptible to the illusion, or do they differ in mobility? If they differ in mobility, what are the factors that account for this difference?

It is evident that the two lights may differ in respect to the *frequency* and *degree* to which they are susceptible to the illusion. Any accurate measure of relative mobility must take account of both factors. In order to do so we have tabulated our data in the following manner. It may be recalled that the subject was required to give a judgment in each case as to the relative amount of motion exhibited by the two lights. Accordingly the judgments were classified on the following basis:

1. Those in which one light, the smaller one for example, was reported as doing all of the moving or moving faster than the second light and hence exhibiting the greater amount of movement.

2. Those in which the second light, the larger one, was reported as doing all of the moving or moving faster than the first light and hence exhibiting the greater amount of movement.

3. Those in which an equal amount of movement was reported.

Inasmuch as the two lights, for all possible comparisons, actually moved an equal number of times, the above values necessarily constitute a measure of their relative mobility. Consequently in all of our tables, the above three percentage values are stated in terms of relative mobility.

Size

Small Light vs. Large Light.

Is the smaller light more mobile than the larger light? Group results are presented in Table XIII.

The smaller light is very much more mobile than the larger light as can be seen from the data on average speed. The

TABLE XIII

SIZE EFFECTS ON MOBILITY

Effects of Speed on Mobility Due to Size

	Aver. Speed	Speed *A*	Speed *B*	Speed *C*
Smaller More Mobile.....	47.37	48.46	45.46	48.17
Larger More Mobile.....	31.23	24.62	31.17	37.92
No Difference.....	21.40	26.92	23.37	13.91

Group results. 7,200 reactions.

differences are significant but reliability figures are not presented because there is perfect agreement among the subjects without exception.

But would increasing the speed of the lights alter the relative mobility to the lights? Table XIII again offers the pertinent data. Our dichotomization shows that as the speed is increased the difference between the perceived mobility of the smaller light and the larger light is lessened. This is due to an increase of judgments which ascribe greater mobility to the larger light. The number of judgments favoring the mobility of the smaller light remains about the same throughout all speeds, while the number of judgments declaring 'no difference' is lessened as speed is increased.

Again it is not necessary to give reliability figures because of consistent agreement among the subjects. One subject, however, showed no effect of speed on mobility due to size.

Difference in Size.

Will the relative mobility of the smaller light in relation to the larger light be affected by a change in difference in size of the two lights from a great difference in size (pair I), to a medium difference in size (pair II), to a small difference in size (pair III)?

TABLE XIV*a*

DIFFERENCE IN SIZE EFFECTS ON MOBILITY

	Pair I	Pair II	Pair III
Smaller More Mobile.....	50.92	46.88	44.29
Larger More Mobile.....	29.04	31.37	33.29
No Difference.....	20.04	21.75	22.42

Group results. 7,200 reactions.

From Table XIV*a* it is apparent not only that there is a decrease of percentage for the smaller light being mobile and an increase of percentage for the larger light being mobile, but that the difference in favor of mobility of the smaller light becomes less and less with a decrease in difference in size.

Eight of the ten subjects show less tendency to regard the smaller light as more mobile when the medium difference in size is contrasted with the greatest difference in size. Nine of the ten subjects show less tendency to regard the smaller light as more mobile when the least difference in size is compared with the medium difference in size. All ten of the subjects show less tendency to regard the smaller light as more mobile when the least difference in size is contrasted with the greatest difference in size. The reliability of group differences is shown in Table XIV*b*.

TABLE XIV*b*

DIFFERENCE IN SIZE EFFECTS ON MOBILITY

Reliability Table

	Pair I	Pair II	Pair III
S Mobile–L Mobile	21.88	15.51	11.00
	Diff.	S.E. Diff.	$\frac{D}{\text{S.E. Diff.}}$
I–II	6.37	1.25	5.3
II–III	4.51	1.26	3.6
I–III	10.88	1.25	8.7

We may conclude that with the majority of individuals a decrease in difference in size between the smaller and larger light decreases the tendency to see the smaller light as relatively more mobile than the larger light.

Fixation

Fixated Light vs. Non-Fixated Light.

Is fixation a factor which may influence mobility? Will fixation of a light tend to make that light more mobile?

Data on average speed indicates that the effect of fixation

TABLE XV

FIXATION EFFECTS ON MOBILITY

Effects of Speed on Mobility Due to Fixation

	Aver. Speed	Speed *A*	Speed *B*	Speed *C*
Fixated More Mobile.....	49.95	51.00	50.58	48.25
Non-Fixated More Mobile..	28.65	22.08	26.04	37.83
No Difference.....	21.40	26.92	23.38	13.92

Group results. 7,200 reactions.

on mobility is pronounced. Group results show that the fixated light has greater mobility than the non-fixated light. This is true for all subjects.

When the speed is increased will the mobility due to fixation be changed? Reliability of group differences have not been included in Table XV because of good individual agreement. The mobility of a fixated light is slightly reduced as speed is increased, but this difference is not reliable. However, the mobility of the non-fixated light does increase with speed and these differences are statistically significant. At the same time there is a pronounced falling off of judgments 'no difference' with increased speed.

Individual records with the exception of one subject, who is unaffected, reveal that with increasing speed the proportion of judgments in favor of the mobility of the fixated becomes less, due to an increase in judgments giving greater mobility to the non-fixated light.

What effect will a greater or less difference in size between two lights have upon the high relative mobility of the fixated light, and the low relative mobility of the non-fixated light? Table XVI gives group results.

TABLE XVI

FIXATION EFFECTS ON MOBILITY

Effect of Difference in Size on Mobility Due to Fixation

	Pair I	Pair II	Pair III
Fixated More Mobile.....	50.75	49.45	49.62
Non-Fixated More Mobile.....	29.21	28.80	27.96
No Difference.....	20.04	21.75	22.42

Group results. 7,200 reactions.

None of the differences for the group are found to be reliable. For all individuals the results are purely negative, *i.e.*, whatever the difference in size the mobility of the fixated light remains the same.

Right Light vs. Left Light.

When the right light is fixated or the left light is fixated, will the difference in this direction of fixation affect the higher mobility of the fixated light and the lower mobility of the non-fixated light? Table XVII contains the data on this question.

TABLE XVII

FIXATION EFFECTS ON MOBILITY

Right Light Fixated vs. Left Light Fixated

	Right Light Fixated	Left Light Fixated
Fixated More Mobile	50.75	49.14
Non-Fixated More Mobile	28.58	28.72
No Difference	20.67	22.14

Group results. 7,200 reactions.

It is apparent that there are no significant differences, individual variations tending to balance each other. An examination of individual records reveals that three subjects, HA, AU, and TA, increase their tendencies to regard the fixated light as mobile through right fixation; three individuals, OK, HO, and ZE by fixating the left light increase the mobility of the fixated light.

Position

What effect will position have upon mobility? Will a light on the right of the observer be more mobile than a light on the left, or vice versa? It was shown that direction of fixation was important for mobility with six of the ten subjects, but individual records were not presented because the greater mobility of a right or left light for any subject can be shown in position effects.

Five subjects are unaffected by position. Greater mobility is ascribed to the left light by ZE and HO. Greater mobility is ascribed to the right light by HA, AU and TA.

TABLE XVIII*a*
POSITION EFFECTS ON MOBILITY

	Right More Mobile	Left More Mobile	No Difference
ZE	21.67	38.89	39.44
GR	28.06	25.14	46.80
CU	39.45	39.45	21.10
HA	56.11	43.89	0
AU	51.25	31.25	17.50
OK	37.08	43.89	19.03
WA	40.56	42.22	17.22
TA	50.28	40.14	9.58
WH	46.39	48.33	5.28
HO	26.39	35.55	38.06
	Individual results.	7,200 reactions.	
GROUP	39.72	38.88	21.40

These differences are reliable as can be seen from Table XVIII*b*.

TABLE XVIII*b*
POSITION EFFECTS ON MOBILITY
Reliability Table

	Difference (Right minus Left)	S.E. Diff.	$\frac{D}{\text{S.E. Diff.}}$
ZE	−17.22	2.38	7.2 (L)
GR	2.92	2.33	1.2
CU	0	—	—
HA	12.22	2.61	4.7 (R)
AU	20.00	2.54	7.8 (R)
OK	− 6.81	2.59	2.6
WA	− 1.66	2.60	.6
TA	10.14	2.61	3.9 (R)
WH	− 1.94	2.63	.7
HO	− 9.16	2.44	3.7 (L)
GROUP	.84	.81	1.0

Relative Strength of the Factors Fixation and Size on Mobility

Irrespective of size, position, speed, or any of the factors we have considered, the fixation of a light makes that light mobile. Likewise a small light has been found to be relatively mobile, a large light relatively less mobile. Now from these two factors with their effects upon mobility it would be expected that when a small light is fixated the mobility of that light would become increasingly greater than when that light is not fixated. Also, the fixation of a large light would lend

mobility to that light which it would not otherwise possess. The strength of these two factors may be analyzed and compared as is done in Table XIX*a*.

TABLE XIX*a*

RELATIVE STRENGTH OF THE FACTORS FIXATION AND SIZE ON MOBILITY

Judgs.	Small Fixated			Large Fixated		
	Large Mobile	Small Mobile	No Diff.	Large Mobile	Small Mobile	No Diff.
ZE.....	15.55	47.78	36.67	22.78	35.00	42.22
GR.....	7.50	55.28	37.22	30.27	13.33	56.40
CU.....	30.00	50.83	19.17	39.45	37.50	23.05
HA.....	28.61	71.39	0	56.67	43.33	0
AU.....	32.22	55.28	12.50	30.83	46.67	22.50
OK.....	31.11	48.06	20.83	35.56	47.22	17.22
WA.....	26.39	57.78	15.83	38.33	43.06	18.61
TA.....	19.45	70.83	9.72	61.39	29.17	9.44
WH.....	18.33	77.50	4.17	65.00	28.61	6.39
HO.....	8.89	57.50	33.61	26.39	33.11	42.50
GROUP..	21.81	59.22	18.97	40.67	35.50	23.83

7,200 reactions.

When the larger light is fixated group results show that the larger light actually is judged more mobile than the smaller light. This is because of the five subjects who are apparently influenced more by fixation than by size. On the other hand, the other five subjects still persist in ascribing more mobility to the smaller light, but in less degree than when the smaller was the fixated light.

TABLE XIX*b*

RELATIVE STRENGTH OF THE FACTORS FIXATION AND SIZE ON MOBILITY

Reliability Table

Larger Light Fixated

	Av. Difference S Mobile—L Mobile	S.E. Diff.	$\frac{D}{\text{S.E. Diff.}}$
ZE, AU, OK, WA, HO.....	9.83	1.59	6.2 (Size)
GR, CU, HA, TA, WH.....	−20.17	1.60	12.6 (Fixation)
GROUP.....	−5.17	1.14	4.5 (Fixation)

Group results indicate that fixation is a stronger factor in affecting mobility than size, but individual records show that half of the subjects are significantly affected by size, half are

significantly affected by fixation when these two factors are made to oppose each other. This, however, is based upon an average of three pairs of sizes.

A complete analysis of the data indicates that when the difference in size decreases the strength of size decreases, or when the difference in size increases the strength of size increases because the larger a light is in relation to a smaller light, the greater the tendency to ascribe more mobility to the small light and less mobility to the large light. This is true for nine of the ten subjects and indicates that the relative strength of size and fixation on relative mobility is dependent upon the difference in size between the two lights.

Summary: Factors Affecting Mobility.

One light is perceived more mobile than the other when:

(1) It is smaller.
(2) It is fixated.

One light is increasingly more mobile when:

(1) It is smaller and the speed is decreased.
(2) It is smaller and the difference in size is increased.
(3) It is fixated and the speed is decreased.
(4) It is both a fixated light and a small light.

There is no change in relative mobility of one light when:

(1) It is fixated and there is an increase or decrease of difference in size.
(2) It is fixated and fixation is varied to the right or to the left.
(3) It is a right light or a left light.

A test of the relative strength of the mobility factors fixation and size shows that fixation is the stronger factor for the group, but the subjects are equally divided in number as to preference of one factor or the other. However, the relative strength of fixation and size is dependent upon the difference in size.

(Manuscript received February 20, 1933)

OVER-COMPENSATION BY THE NON-PREFERRED HAND IN AN ACTION-CURRENT STUDY OF SIMULTANEOUS MOVEMENTS OF THE FINGERS

BY MILTON METFESSEL AND NEIL D. WARREN

University of Southern California

Originally the experiments here outlined were intended to compare the relative handedness-indicating value of prior action currents and prior movements in simultaneous flexion of the first two fingers of each hand. For this purpose a total of 88 subjects was divided into four groups: 23 dextrals, 22 sinistrals, 21 dextro-sinistrals, and 24 stutterers. From these we secured 1906 measurable action current reactions and 2555 movement records. The justification for the grouping of stutterers in a separate class lay in a large body of experimental evidence indicating a relationship between stuttering and native left-handedness. (1, 3, 4.) In this group of stutterers there were two left-handed subjects who were classified also in the sinistral group. The sinistral group included only those who use the left hand in writing. In view of the social pressure to use the right hand in writing, persistence in the use of the left is regarded as the most significant indicator of sinistrality. All subjects who reported a history of change of handedness from left to right, and those who use the left hand for most performances but write with the right hand were classified as dextro-sinistrals. The age of the subjects varied from 14 to 57 years with 85% between the ages of 18 to 25 years.

APPARATUS

The complete apparatus consisted essentially of devices for picking up, amplifying, and recording the action currents from the two forearms during simultaneous voluntary flexion of the first two fingers of each hand.[1] It had the advantage of giving at the

[1] The authors acknowledge gratefully the assistance of Mr. W. F. Wichart in the construction of the apparatus.

same time a record of the effective movement of each hand. The devices for the two hands were duplicates with the end result being recorded upon a single strip of photographic paper. A further advantage of this apparatus was its economy of operation, since 24 reactions for both hands could be photographed upon one strip of paper 1.3 meters in length.

Electrodes and Movement Keys.—Two electrodes were attached to the inner surface of the forearms at a distance of 20 mm to 100 mm apart. The lower electrode was placed at the base of the palm of the hand and the upper one was attached at a distance which gave a response amplitude of from 2 mm to 4 mm. The electrodes were flexible copper strips 25 mm by 75 mm in area. Small hooks at the end of each electrode permitted its attachment by means of rubber bands. A felt pad, 5 mm thick, slightly larger than the metal strip, was moistened in a saturated salt solution and placed between the electrode and the skin. Wires led from each electrode to the amplifiers.

A telegraph key was placed in the circuit from the lower electrode to the amplifier. The key was so adjusted that a slight pressure would break the action-current circuit, thereby making a record of the effective movement of the fingers. The keys were of equivalent tension, 290 grams being required to break the circuit for the right hand key and 295 grams for the left. They were mounted on heavy wooden blocks on the side away from the subject, so that movement of the fingers in pressing the keys was toward the subject's body. The forearms rested on Kapok-filled pads which were intended to reduce implicit extraneous movements during the tests.

Amplifiers.—The amplifiers used were of conventional design involving three stages of resistance-coupled amplification. They were sensitive to external electrical phenomena, such as commercial radio waves and the radiations from the lighting circuit of the building. To overcome this difficulty the entire set-up was enclosed in a metal shield: a box, 6 feet high, 6 feet wide, and 12 feet long. The shield served also as a dark room.

Recording Camera.—Two Westinghouse oscillograph units were mounted, one above the other, so as to reflect two beams of light through a lens system onto the surface of a revolving drum. (See Fig. 1.) The oscillographs were activated by the action currents through the amplifiers. The vibrations of the reflecting mirror were photographically recorded on sensitized paper. A strip of this paper (Eastman Press Bromide) 35 mm wide was attached to the drum, which was 415 mm in diameter and 50 mm wide. A projecting rim 5 mm from the top served as a guide in adjusting the sensitized paper. A simple method of holding the paper in place proved convenient. After it was placed around the drum the ends were inserted through a narrow slot in the drum and caught inside by a heavy perforating clip. An anchored spring pulled the clip toward the center of the drum, furnishing enough tension to prevent the paper from slipping, at the same time permitting easy alignment of the paper.

The drum was made of light weight aluminum and had little inertia. It was revolved horizontally by a phonograph motor at a speed of 45 r.p.m. At this speed 1 mm of the recording paper passed a given point in .001 second. An adjusting screw moved each oscillograph mirror to two positions on the upper half of the sensitized paper. After this portion had been exposed the paper was turned over on the drum so as to expose the other half. By this method it was possible to obtain a recording at four different vertical positions on the paper.

Six records were made on each circumference of the paper so that a total of 24 trials were recorded on each strip. This was accomplished by means of an automatic magnetic shutter in front of the light source which exposed the light for only one-sixth of a revolution. A sliding contact on the drum closed the circuit which operated the

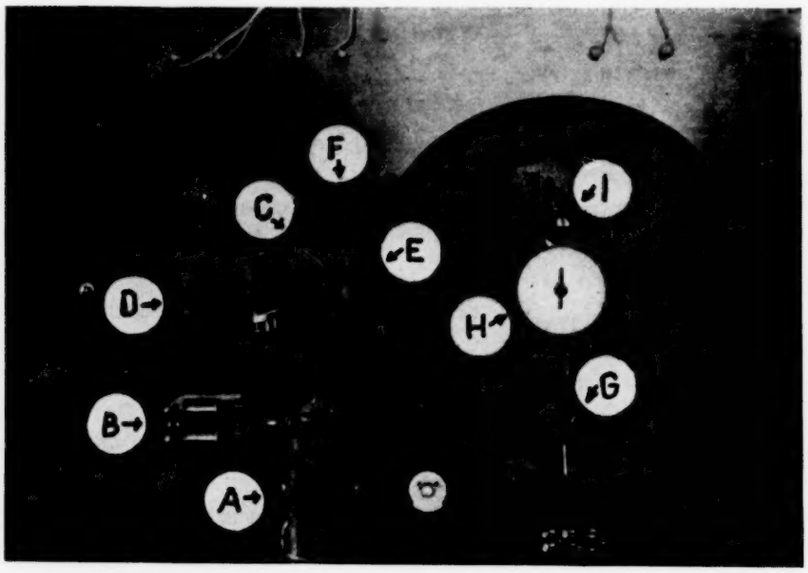


FIG. 1. View of photographic unit from above. *A*. Light source. *B*. Magnetic shutter. *C*. Mirror for light reflections. *D*. Oscillographs. *E*. Lens. *F*. Shield for adjustment of lights to two positions on the drum. It is flipped in front of the lens for this operation. *G*. Anchored perforating spring clip for holding photographic paper in place. *H*. Stroboscopic disc for timing. *I*. Adjusting Knob for sliding contacts operating magnetic shutter (*B*) for one-sixth of a revolution, and also controlling the signal light. (*C* in Fig. 2.) The sliding contacts are under the drum.

automatic shutter and, four hundredths of a second later, lighted a faint green signal lamp in front of the subject. The response must have been made within a period of .217 second, or .177 second after the signal light appeared, in order to record. The experimenter closed a switch for one selected revolution of the drum so as to allow one exposure of each area of the recording paper. After each trial an adjusting knob on the drum set the contact to expose the next segment.

PROCEDURE

The subject was seated comfortably before the table upon which the two telegraph keys were mounted. (See Fig. 2.) His position was adjusted so he could easily reach the keys. The electrodes were attached and the subject clenched his hands to permit the experimenter to note the oscillograph response. Any necessary adjustments in electrode distance were made to secure approximately equal action-current amplitudes from the two arms.

The subject placed the first two fingers of each hand lightly upon the surface of the keys. The green signal light was turned on and the subject was instructed to press both keys simultaneously with each flash of the light. He was requested to relax as much as possible between signals so as to prevent any extraneous movement, and to make each pressure a sudden, sharp movement.

Three to five practice series were given before the sensitized paper was placed on the drum. The subject was told when to begin and was asked to continue pressing the keys as long as the light continued flashing. Most subjects adapted rapidly to the rhythm of the performance so that no further instruction or practice was required. In the actual test the experimenter waited until the subject had secured the rhythm before

exposing the sensitized paper. After the first few trials the shutter was opened on the third or fourth response. Each recorded response constituted a single trial. The earlier subjects were given 24 trials but the majority (59 percent) were given 48 trials. The administration of the tests required an average of about 20 minutes. Brief rest periods were permitted at the end of each twelve trials.

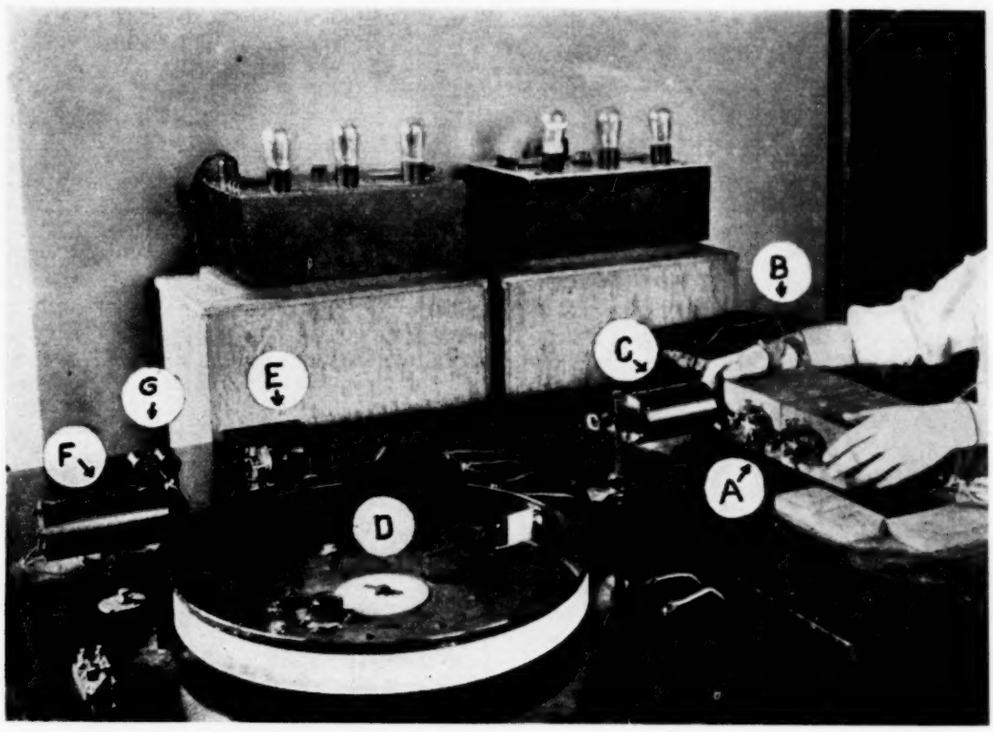


FIG. 2. Showing photographic unit, amplifiers, and reaction unit. *A*. Telegraph keys. *B*. Electrodes. These were moved upward slightly for the photograph. *C*. Faint green light stimulus. *D*. Photographic drum. *E*. Oscillographs. *F*. Light source. *G*. Magnetic shutter, operated by sliding contact under photographic drum, *D*.

MEASUREMENT OF RECORDS

The photographed result of each trial was a pair of lines, the upper for the right hand and the lower for the left (Fig. 3). The actual movement of the keys produced a break in the lines while the action-currents preceding the movement appeared as verti-

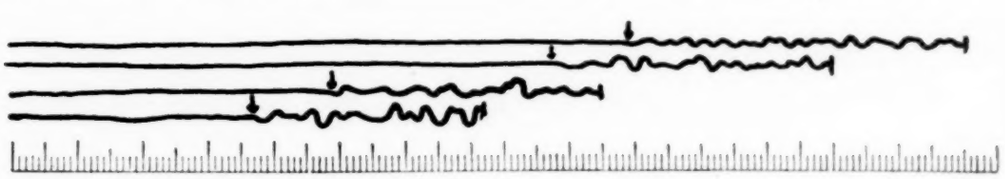


FIG. 3. Showing two trials of a right-handed subject. Upper line in each pair is the right hand, and lower the left. Action current starts are indicated by arrows. The movement is shown by the cessation of the photographic line. Scale in mm. or ms.

cal waves. In order to indicate the starts of the action currents the beginning of a continuous series of waves immediately preceding the actual movement was marked in each trial. Any earlier waves, not continuous with this series, were considered in this

experiment as evidence of movements not involved in actually pressing the keys. The extent of the lead of one hand over the other, in both action currents and movement were recorded in terms of milliseconds (ms).

In Tables I to IV the average leads for both action currents and effective movements for the subjects are listed by groups. The reliability of each average is indicated, also, in terms of the number of chances in 100 that the average will always favor the hand indicated. The averages give the number of milliseconds by which the hand indicated preceded the other.

TABLE I

DEXTRALS

Subject Number	Action Currents				Movement			
	Leading Hand	Average (ms)	σ Av. (ms)	% Chances	Leading Hand	Average (ms)	σ Av. (ms)	% Chances
1	Left	12.3	±2.2	100	Left	1.8	±1.3	91
2	Left	6.8	±1.3	100	Right	3.1	±0.8	100
3	Right	0.8	±2.2	64	Right	2.6	±1.6	95
4	Right	11.1	±1.7	100	Right	1.7	±1.2	92
5	Right	1.5	±2.3	74	Left	4.4	±2.5	96
6	Left	12.5	±3.3	100	Left	9.8	±2.1	100
7	Left	11.1	±2.1	100	Left	18.3	±1.6	100
8	Left	7.7	±1.2	100	Left	5.3	±1.6	100
9	Left	10.6	±1.9	100	Left	1.2	±0.7	91
10	Right	0.3	±3.5	54	Right	11.4	±2.4	100
11	Left	20.7	±2.6	100	Left	14.8	±1.7	100
12					Left	7.1	±4.0	96
13	Right	1.8	±1.1	94	Left	3.3	±1.1	100
14					Left	1.6	±2.6	74
15	Left	15.3	±2.8	100	Left	9.3	±2.4	100
16	Right	11.5	±2.0	100	Right	12.4	±1.3	100
17	Right	3.9	±1.6	99	Left	3.4	±1.6	98
18	Left	14.4	±1.9	100	Left	14.3	±2.0	100
19	Left	4.6	±3.1	93	Left	4.8	±2.7	96
20	Right	4.7	±4.1	87	Left	3.3	±2.7	89
21	Right	2.6	±1.5	96	Left	1.8	±2.0	82
22	Right	13.9	±0.8	100	Left	10.3	±1.0	100
23	Right	5.4	±1.8	100	Right	8.6	±1.3	100
Group Average	Left	3.1	±0.7	100	Left	3.4	±0.6	100

RESULTS FOR ACTION-CURRENT LEADS

The results for action-current leads are found to have a bearing on the following problems: (1) the difference between the subjects in the various classifications, and (2) the value of action-current leads in indicating a unilateral dominance.

In Table V it can be seen that no reliable difference in action-current lead averages is found between the dextral group and the dextro-sinistral group, nor between these groups and stutterers. The sinistral group, however, varies from dextrals

TABLE II

SINISTRALS

Subject Number	Action Currents				Movement			
	Leading Hand	Average (ms)	σ Av. (ms)	% Chances	Leading Hand	Average (ms)	σ Av. (ms)	% Chances
1	Right	4.2	±1.8	99	Right	6.3	±1.1	100
2	Right	7.6	±2.2	100	Left	5.0	±1.7	100
3[2]	Right	9.7	±2.4	100	Left	9.1	±1.5	100
4	Left	3.2	±2.0	89	Left	2.8	±1.1	99
5	Right	16.7	±2.5	100	Right	10.3	±1.9	100
6	Right	11.4	±1.0	100	Right	8.6	±1.6	100
7	Left	7.3	±1.4	100	Left	9.4	±1.0	100
8	Left	0.1	±1.4	52		0.0	±2.0	
9	Right	11.0	±3.5	100	Right	4.7	±1.6	100
10	Right	22.1	±2.2	100	Right	11.9	±1.8	100
11	Right	7.1	±1.9	100	Right	2.1	±0.8	100
12	Right	3.4	±1.9	97	Right	3.5	±1.4	99
13	Right	10.6	±3.6	100	Right	4.2	±3.8	87
14	Right	13.0	±6.6	93	Left	2.4	±3.4	76
15	Right	0.4	±3.4	54	Left	0.7	±2.1	63
16[2]	Right	7.9	±1.6	100	Right	1.3	±2.1	74
17	Right	15.5	±2.9	100	Right	17.2	±2.3	100
18	Right	0.2	±1.7	55	Left	7.3	±1.3	100
19	Left	9.0	±1.0	100	Left	13.9	±1.2	100
20	Left	12.6	±2.3	100	Left	7.7	±1.6	100
21	Left	3.5	±2.0	96	Left	4.2	±2.1	96
22	Left	8.5	±3.8	99	Left	18.0	±2.3	100
Group Average	Right	3.4	±0.7	100	Left	0.5	±0.5	83

by 6.6±1.0 ms, from dextro-sinistrals by 7.8±0.8 ms, and from stutterers by 6.9±1.1 ms. Each of these differences is reliable in light of its standard error.

The individual results, as listed in Tables I to IV, indicate the same thing. In Table I those action-current averages which are reliably greater than zero (100 percent) are divided 9 left and 4 right. Left-handed individuals (Table II), on the other hand, show 10 reliable right averages and 3 left. Dextro-sinistrals (Table III) again resemble dextrals since the reliable lead averages are 10 left and 1 right. The same thing is true for stutterers (Table IV) with 6 reliable left averages and 4 right.[3] It is evident, therefore, that in action-current leads, the stutterers and the dextro-sinistral group resembled the dextrals rather than the sinistrals.

[2] Subjects 3 and 16 in this list are left-handed stutterers.

[3] Two of the reliable right-hand averages were made by left-handed stutterers.

TABLE III

DEXTRO-SINISTRALS

Subject Number	Action Currents				Movement			
	Leading Hand	Average (ms)	σ Av. (ms)	% Chances	Leading Hand	Average (ms)	σ Av. (ms)	% Chances
1	Right	3.3	±1.7	97	Right	8.8	±1.1	100
2	Left	22.3	±5.9	100	Left	16.2	±2.2	100
3	Left	6.1	±1.6	100	Right	5.5	±1.7	100
4	Right	1.2	±1.7	73	Right	6.9	±2.1	100
5	Left	16.6	±1.8	100	Left	8.9	±1.1	100
6	Right	0.9	±1.2	77	Right	2.9	±0.8	100
7	Left	0.3	±1.7	59	Left	3.3	±2.3	93
8	Left	8.7	±1.2	100	Left	5.3	±1.2	100
9	Left	10.0	±3.0	100	Left	4.8	±1.9	99
10	Left	1.2	±2.1	71	Left	8.5	±1.7	100
11	Right	10.5	±2.8	100	Right	3.7	±3.9	83
12	Left	6.3	±1.5	100	Left	8.9	±1.0	100
13	Left	2.3	±1.7	92	Left	2.4	±1.1	99
14	Right	2.3	±1.8	90	Left	4.4	±1.5	100
15	Left	13.3	±2.3	100	Left	7.6	±1.0	100
16	Left	9.3	±2.2	100	Left	3.3	±1.5	98
17	Left	9.9	±1.6	100	Left	17.9	±3.2	100
18	Left	2.3	±1.5	74	Left	2.6	±1.5	96
19	Left	14.4	±1.7	100	Left	11.1	±2.1	100
20	Right	2.9	±1.2	99	Left	0.8	±1.0	78
21	Left	3.3	±2.1	95	Right	0.6	±1.7	65
Group Average	Left	4.4	±0.5	100	Left	3.4	±0.6	100

That prior action-current leads in a hand under these conditions are not indicative of native physiological preference for handedness is apparent in both group and individual results. (Contrast **3**, **5**.) The group averages for dextrals, dextro-sinistrals, and stutterers, show a definite left-hand lead, while the average for the left-handed group indicates an equally definite right-hand advantage. Of all the right-handed writers (dextrals, dextro-sinistrals, and stutterers) having reliable action-current lead averages, 78.1 percent have left leads. For left-handed writers, 76.9 percent of reliable leads indicate the right hand. Thus, a definite tendency toward an action-current lead by the non-preferred hand is evident in both group results and individual averages.

RESULTS FOR MOVEMENT LEADS

Essentially the same results are found in movement records as in action currents. In group averages (Table V)

TABLE IV

STUTTERERS

Subject Number	Action Currents				Movement			
	Leading Hand	Average (ms)	σ Av. (ms)	% Chances	Leading Hand	Average (ms)	σ Av. (ms)	% Chances
1	Left	5.2	±1.4	100	Left	0.6	±0.4	95
2	Left	8.3	±2.9	100	Left	6.7	±1.5	100
3	Left	22.0	±4.0	100	Left	21.7	±3.3	100
4	Left	2.9	±1.2	99	Left	8.3	±1.2	100
5	Left	18.0	±2.2	100	Left	8.2	±1.2	100
6	Left	4.7	±2.6	97	Left	1.3	±2.0	100
7[4]	Right	9.7	±2.4	100	Left	9.1	±1.5	100
8	Right	4.2	±3.9	86	Right	7.2	±1.9	100
9					Left	9.8	±4.5	99
10	Left	1.3	±3.2	65	Right	11.6	±5.7	98
11	Right	16.4	±3.6	100	Left	0.8	±4.7	57
12					Left	9.0	±1.4	100
13	Left	6.3	±3.3	97	Left	11.5	±1.6	100
14	Left	2.3	±3.2	76	Left	4.8	±1.9	99
15	Right	1.3	±3.0	67	Left	14.1	±5.1	100
16	Left	20.0	±5.3	100	Left	13.5	±4.7	100
17					Left	12.9	±1.7	100
18	Left	1.4	±2.7	69	Left	4.9	±1.3	100
19	Right	8.5	±1.5	100	Right	4.7	±1.6	100
20[3]	Right	7.9	±1.6	100	Right	1.3	±2.1	74
21	Right	0.8	±1.7	70	Left	6.7	±1.4	100
22	Left	13.8	±2.2	100	Left	10.7	±2.3	100
23					Left	11.0	±2.7	100
Group Average	Left	3.4	±0.8	100	Left	7.4	±0.6	100

[4] Subjects 7 and 20 in this group are left-handed stutterers.

the dextrals and dextro-sinistrals are not significantly different. Both the stutterers and the sinistrals differ from these groups by reliable amounts but toward the opposite extremes. The stutterers show an average movement lead indicating a left-hand precedence of 7.4±0.6 ms. The same hand has the advantage as in the averages of the dextrals and dextro-sinistrals with the average lead difference being more pronounced. The sinistral group average is not reliably greater than zero. While it indicates a left-hand advantage of 0.5±0.5 ms, the true average may lie between a right-hand lead of 1.0 ms and a left-hand lead of 2.0 ms. The difference found between the dextrals and the sinistrals, 2.9±0.8 ms, is reliable in terms of its standard error.

The number of reliable movement averages for individuals

TABLE V

DIFFERENCES IN GROUP AVERAGES

Group	Action Currents			Movement		
	Difference between Averages (ms)	σ Diff. (ms)	% Chances of Reliable Difference	Difference between Averages (ms)	σ Diff. (ms)	% Chances of Reliable Difference
Dextrals and Sinistrals	6.5	±1.0	100	2.9	±0.8	100
Dextrals and Dextro-sinistrals	1.3	±0.9	93	0.6	±0.8	77
Dextrals and Stutterers	0.4	±1.1	63	4.0	±0.8	100
Dextro-sinistrals and Sinistrals	7.8	±0.8	100	3.4	±0.7	100
Dextro-sinistrals and Stutterers	0.9	±0.9	84	3.4	±0.7	100
Sinistrals and Stutterers	6.9	±1.1	100	6.9	±0.8	100

in the various groups is as follows: dextrals, 9 left and 4 right; sinistrals, 7 left and 7 right; dextro-sinistrals, 9 left and 4 right; stutterers, 15 left and 2 right.[5]

The group averages and individual results for movement leads indicate that the dextrals and the dextro-sinistrals are nearly identical in the precedence of the left hand. The stutterers resemble these groups more than the sinistral group. (Contrast **2**, **5**.) Of all the right-handed writers having movement-lead averages reliably greater than zero, 75.6 percent indicate a left-hand advantage.[6] The left-handed writers show no difference in individual movement leads, the advantage being equally divided between the two hands.

COMPARISON OF ACTION-CURRENT AND MOVEMENT LEADS

A tabulation of all trials in which both action-current and movement leads could be measured shows a close correspondence between the two measures. The action currents appear first in the arm which moves first 79.2±1.0 percent of the time. The correspondence between the extents of action-

[5] One left-hand average was the result of a left-handed stutterer.

[6] At the time of proofreading this material, additional data can be reported. A simple method was devised to indicate the prior movement of the fingers of the right or left hand by means of the flash of a neon lamp. A test of 46 right-handed normal speakers produced the following results: 30 cases had consistent left hand leads; 12 had consistent right hand leads; in 4 cases neither hand had the advantage.

current leads and movement leads is indicated by a coefficient of correlation of $+.59\pm.01$.

A measurement of the action-current leads in those trials in which the movement was simultaneous results in averages not reliably greater than zero (Table VI). For these measurements all movements within three ms of each other are considered simultaneous. This figure was selected arbitrarily with some justification in the fact that there were only two reliable averages of three ms or less in the individual cases.

TABLE VI

ACTION-CURRENT LEADS IN SIMULTANEOUS MOVEMENTS

Group	Action Currents				% of Total Responses within 3 ms
	Leading Hand	Average (ms)	σ Av. (ms)	% Chances	
Dextrals	Left	0.8	±1.1	76	27.1
Sinistrals	Right	1.9	±0.8	99	31.7
Dextro-sinistrals	Left	2.4	±1.8	91	32.8
Stutterers	Right	0.2	±1.6	56	22.2

A significant difference between normal speakers and stutterers in the number of measurable records indicates a lesser degree of ability to respond as required by the experimental procedure on the part of the stutterers. The percentage of total trials which could be measured is given in Table VII. The difference of 16.8 ± 3.0 percent of measurable action-current records is reliable, while the difference of 4.4 ± 2.2 percent in movement records has a reliability of 98 chances in 100. The action-current record was not measurable in cases where the subject failed to relax between responses. In in-

TABLE VII

PERCENTAGE OF MEASURABLE TRIALS FOR NORMAL SPEAKERS AND STUTTERERS

Group	Action Currents		Movement	
	%	σ%	%	σ%
Normal Speakers	59.3	±1.3	73.6	±1.0
Stutterers	42.5	±2.7	69.2	±2.0
Difference	16.8	±3.0	4.4	±2.2

stances when the subjects did not react in rhythm to the light stimulus, neither action-current nor movement leads could be determined.

Conclusions

1. Attempted simultaneous movements of the first two fingers of each hand began simultaneously 26.8 percent of the time. 'Simultaneous' is defined to include all movements within three milliseconds of each other.

2. In the case of simultaneous movements, the average action-current lead did not significantly favor either hand.

3. Since there was a correlation of $+.59 \pm .01$ between the extents of action-current leads and movement leads, and since the two measures indicated the same hand in 79.2 ± 1.0 percent of the trials, movement leads seem as significant as action-currents in relation to handedness.

4. In both action-current leads and movement leads, stutterers and dextro-sinistrals resembled dextrals more than they did sinistrals.

5. There was a significant difference between dextrals and sinistrals in both action-current leads and movement leads.

6. Taken as a whole, there was a definite tendency for both action-current leads and movement leads to favor the hand not used in writing, probably due to over-compensation of the less-used hand in an attempt at simultaneous movement.

(Manuscript received January 31, 1934)

References

1. Oates, D. W., Left-handedness in relation to speech defects, intelligence, and achievement. *Forum of Educ.*, 1929, **7**, 91–103.
2. Orton, S. T. and Travis, L. E., Studies in stuttering: Studies of action currents in stutterers, *Arch. Neur. and Psychiat.*, 1929, **21**, 61–68.
3. Travis, L. E., *Speech Pathology*, New York, D. Appleton and Co., 1931, pp. 331.
4. Travis, L. E., Recurrence of stuttering following shift from normal to mirror writing, *Arch. Neur. and Psychiat.*, 1929, **21**, 386–391.
5. Travis, L. E., and Herren, R. Y., Studies in stuttering; *V*, A study of simultaneous antitropic movements of the hands of stutterers, *Arch. Neur. and Psychiat.*, 1929, **21**, 487–494.

THE EFFECT OF FATIGUE UPON VARIABILITY OF OUTPUT IN MUSCULAR WORK

BY CHARLES W. MANZER

New York University

Statement of the Problem

This paper reports an investigation of the changes in absolute and in relative variability of muscular work which accompany the work decrement.

The most extensive study of variability in muscular work is that by Weinland (**4**), who found that fatigue was accompanied by large increases in variability. Changes in variability due to learning have been studied recently by Asch (**1**) and by McGeoch (**2**). Both of these investigators found that there is no regular change in variability accompanying learning. The effect of sheer repetition upon variability has been examined by Thorndike (**3**). In his experiment of drawing 3,810 lines with eyes closed, trying to make each line two inches long, Thorndike found that the variability around the subject's own central tendency was reduced from .11 inch to .08 inch when the first 193 responses were compared with the last 202 responses.

Subjects and Procedure

The present data consist of the work done by twenty-seven men college students, each of whom performed eleven periods of work, a total of two hundred ninety-seven work periods. The muscle group used comprised the flexors of the middle finger of the preferred hand; the apparatus was the Mosso ergograph. Contractions were made at a two-second rhythm, and in every period work was continued to exhaustion. The load lifted was a ten-pound weight. A five-minute rest was taken between successive work periods. Contractions were recorded kymographically and measured in units of one twentieth of an inch.

Results

Total Work.—Table I*a* shows the number of muscular contractions, the mean output per contraction, the standard deviation and the coefficient of variation of the total amount of work done by all subjects in each of the eleven work periods. The height through which the ten-pound load was lifted at each contraction was measured (as has been said) to the nearest twentieth of an inch and this unit has been retained in all the computations of the present study.

An examination of the standard deviations and of the coefficients of variation in Table I*a* discloses the fact that there is no systematic change in either absolute or relative variability throughout the series of work periods. In general, the standard deviations decrease as the mean output decreases. The first work period has the largest standard deviation, the second work period has the next largest. The critical ratio (Difference/P.E. $_{\text{Difference}}$) of the difference between these two standard deviations is 4.97. This ratio shows that the absolute variability of the second work period is significantly less than that of the first work period. The third highest standard deviation is found in the last work period. This final increase in absolute variability is possibly due to differences in incentive among the subjects rather than to differences in ability to do work. The standard deviations in Table I*a* justify the statement that in the present work series initial absolute variability is high, this variability decreases sharply in the early work periods, paralleling the work decrement, and fluctuates irregularly throughout the remaining periods of the series.

In the present series, since there are large decreases in mean output in successive work periods, the coefficient of variation is superior to the standard deviation as a measure of variability.[1] Table I*b* contains the differences and the critical ratios of the differences between the coefficient of variation of the first work period and the coefficient of variation of each of the following work periods. It will be seen in

[1] Karl J. Holzinger, *Statistical Methods for Students in Education.* New York: Ginn and Company, 1928, p. 117.

TABLE I*a*

MEAN OUTPUT, STANDARD DEVIATION AND COEFFICIENT OF VARIATION OF THE TOTAL WORK DONE IN EACH OF THE ELEVEN WORK PERIODS
(Number of Subjects = 27)

	Number of Work Period					
	1	2	3	4	5	6
No. of Contractions	1451	1443	1311	1159	1097	864
Mean	16.26 ± .19	14.63 ± .17	13.24 ± .17	12.66 ± .17	11.82 ± .18	11.41 ± .20
Standard Deviation	10.64 ± .13	9.76 ± .12	9.27 ± .12	8.53 ± .12	8.90 ± .13	8.66 ± .14
Coeff. of Variation	65.44 ± .82	66.71 ± .84	70.02 ± .92	67.38 ± .94	75.29 ± 1.08	75.90 ± 1.23

	Number of Work Period				
	7	8	9	10	11
No. of Contractions	736	656	575	541	584
Mean	11.28 ± .22	11.77 ± .23	12.28 ± .25	12.16 ± .25	11.56 ± .26
Standard Deviation	8.67 ± .15	8.73 ± .16	8.74 ± .17	8.48 ± .17	9.47 ± .19
Coeff. of Variation	76.86 ± 1.35	74.17 ± 1.38	71.17 ± 1.42	69.74 ± 1.43	81.92 ± 1.62

Note: The units in this table are twentieths of an inch through which the ten-pound weight was raised.

Table I*b* that all these differences are positive, that is, each of the ten coefficients of variation is greater than that of the first work period. If a critical ratio of four is taken as the criterion of statistical significance, this criterion is not fully satisfied until the fourth work period of the series is reached. Further, the critical ratios at the ninth and the tenth work periods fall below full statistical significance. The trend of these critical ratios shows that relative variability tends to be greatest in the middle work periods of the series, if the final period be excepted for reasons suggested above. The data in Tables I*a* and I*b* seem to justify the conclusion that neither the absolute variability nor the relative variability of total work periods shows any systematic change throughout these work series.

In addition to an examination of the variability of output in the total work periods of the series, two further examinations of variability have been made. The first of these is concerned with the variability of work done with unfatigued muscles. The data of the former analysis consist of the work done in the first ten contractions of each work period; the data of the latter analysis comprise the work done in the last ten contractions of each work period. In order to eliminate the possibility of overlapping between these two groups of ten contractions each, the work of no subject was included in these comparisons who did fewer than six work periods of twenty-five contractions each; also, no work period was included which contained fewer than twenty-five contractions.

Work with Unfatigued Muscles.—Table II*a* shows the number of subjects, the number of contractions, the mean output, the standard deviation and the coefficient of variation of the work done with unfatigued muscles, that is, of the work done in the first ten contractions of each work period. The successive means show a moderate work decrement, but there is no systematic change in the standard deviation from period to period. In the series of total work periods it was found that the first and the last standard deviations were the two largest. Here also, in work done with unfatigued muscles, the first and the last standard deviations are the two largest,

TABLE I*b*

THE DIFFERENCES AND THE CRITICAL RATIOS OF THE DIFFERENCES BETWEEN THE COEFFICIENT OF VARIATION OF THE FIRST WORK PERIOD AND THE COEFFICIENT OF VARIATION OF EACH OF THE FOLLOWING WORK PERIODS

	Work Periods Compared									
	1–2	1–3	1–4	1–5	1–6	1–7	1–8	1–9	1–10	1–11
Difference	+1.27	+4.58	+1.94	+9.85	+10.46	+11.42	+8.73	+5.73	+4.30	+16.48
Diff./P.E.$_{diff.}$	+1.08	+3.72	+1.56	+7.27	+ 7.08	+ 7.23	+5.44	+3.49	+2.61	+ 9.08

Note: The plus sign means that the coefficient of variation of a given work period is greater than that of the first work period.

TABLE II*a*

MEAN OUTPUT, STANDARD DEVIATION AND COEFFICIENT OF VARIATION OF WORK DONE WITH UNFATIGUED MUSCLES (FIRST TEN CONTRACTIONS) IN EACH OF THE ELEVEN WORK PERIODS

	Number of Work Period					
	1	2	3	4	5	6
No. of Subjects	17	17	17	17	17	17
No. of Contractions	170	170	170	170	170	170
Mean	29.55 ± .32	26.55 ± .26	24.20 ± .24	21.97 ± .24	22.20 ± .21	20.91 ± .30
Standard Deviation	6.22 ± .23	5.07 ± .18	4.53 ± .17	4.69 ± .17	4.10 ± .15	5.71 ± .21
Coeff. of Variation	21.05 ± .80	19.10 ± .72	18.72 ± .71	21.35 ± .82	18.47 ± .70	27.31 ± 1.07

	Number of Work Period				
	7	8	9	10	11
No. of Subjects	17	17	15	15	14
No. of Contractions	170	170	150	150	140
Mean	20.57 ± .30	21.00 ± .31	21.11 ± .31	20.11 ± .32	21.43 ± .38
Standard Deviation	5.92 ± .21	4.98 ± .22	5.66 ± .22	5.74 ± .23	6.79 ± .27
Coeff. of Variation	28.78 ± 1.14	23.71 ± .92	26.81 ± 1.05	28.54 ± 1.13	31.68 ± 1.27

Note: The units in this table are twentieths of an inch through which the ten-pound weight was raised.

while the other standard deviations fluctuate irregularly. The present data indicate that work done with unfatigued muscles shows no systematic change in absolute variability.

Table II*b* contains the differences and the critical ratios of the differences between the coefficient of variation of the work done in the first work period with unfatigued muscles and the coefficient of variation of the corresponding work in each of the following work periods. Eight of the ten differences shown in Table II*b* are positive, that is to say, the relative variability of eight of the ten work periods is greater than that of the first period. Five of these eight differences are statistically significant; all of the five significant differences come at the sixth work period or later in the series. The present data justify the statement that, when the comparison is confined to work done with unfatigued muscles, the work done in the later periods of these series is significantly more variable than that done in the earlier periods.

Work with Fatigued Muscles.—In Table III*a* are the number of subjects, the number of contractions, the mean output, the standard deviation and the coefficient of variation of the work done with fatigued muscles, that is to say, of the work done in the last ten contractions of each work period. The successive means show no work decrement. This absence of a decrement is due to the fact that, since the muscles were approaching complete exhaustion, there was little if any reserve energy to expend. The successive standard deviations show no systematic change, but merely fluctuate irregularly. Table III*b* contains the differences and the critical ratios of the differences between the coefficient of variation of the work done in the first work period with fatigued muscles and the coefficient of variation of the corresponding work in each of the following work periods. Eight of these ten differences are positive, but none is statistically significant. The conclusion may be drawn, accordingly, that there is no significant change in relative variability of work performed with fatigued muscles throughout this series of work periods.

The foregoing discussion can be summarized as follows: With the possible exception of work performed with unfa-

TABLE II*b*

THE DIFFERENCES AND THE CRITICAL RATIOS OF THE DIFFERENCES BETWEEN THE COEFFICIENT OF VARIATION OF THE WORK DONE IN THE FIRST WORK PERIOD WITH UNFATIGUED MUSCLES (FIRST TEN CONTRACTIONS) AND THE COEFFICIENT OF VARIATION OF THE CORRESPONDING WORK IN EACH OF THE FOLLOWING WORK PERIODS

	Work Periods Compared									
	1–2	1–3	1–4	1–5	1–6	1–7	1–8	1–9	1–10	1–11
Difference	−1.95	−2.33	+0.30	−2.58	+6.26	+7.73	+2.66	+5.76	+7.49	+10.63
Diff./P.E.$_{diff.}$	−1.80	−2.18	+0.26	−2.43	+4.67	+5.56	+2.18	+4.36	+5.43	+ 7.09

Note: The plus sign means that the coefficient of variation of a given work period is greater than that of the first work period. The minus sign means that the coefficient of variation of a given work period is less than that of the first work period.

TABLE III*a*

MEAN OUTPUT, STANDARD DEVIATION AND COEFFICIENT OF VARIATION OF WORK DONE WITH FATIGUED MUSCLES (LAST TEN CONTRACTIONS) IN EACH OF THE ELEVEN WORK PERIODS

	Number of Work Period					
	1	2	3	4	5	6
No. of Subjects	17	17	17	17	17	17
No. of Contractions	170	170	170	170	170	170
Mean	4.68 ± .24	4.68 ± .24	4.64 ± .24	4.80 ± .25	4.89 ± .25	4.35 ± .22
Standard Deviation	3.18 ± .17	2.99 ± .17	3.20 ± .17	2.97 ± .18	3.38 ± .18	3.34 ± .16
Coeff. of Variation	67.95 ± 3.45	68.98 ± 3.15	68.97 ± 3.52	61.88 ± 3.01	69.12 ± 3.54	76.78 ± 4.47

	Number of Work Period				
	7	8	9	10	11
No. of Subjects	17	17	15	15	13
No. of Contractions	167	167	150	150	130
Mean	5.04 ± .26	4.38 ± .23	4.23 ± .17	4.86 ± .22	4.14 ± .24
Standard Deviation	4.12 ± .18	3.35 ± .16	3.10 ± .12	3.94 ± .16	3.15 ± .17
Coeff. of Variation	81.75 ± 4.89	76.48 ± 4.40	73.29 ± 4.11	81.07 ± 4.80	76.09 ± 4.68

Note: The units in this table are twentieths of an inch through which the ten-pound weight was raised.

TABLE III*b*

THE DIFFERENCES AND THE CRITICAL RATIOS OF THE DIFFERENCES BETWEEN THE COEFFICIENT OF VARIATION OF THE WORK DONE IN THE FIRST WORK PERIOD WITH FATIGUED MUSCLES (LAST TEN CONTRACTIONS) AND THE COEFFICIENT OF VARIATION OF THE CORRESPONDING WORK IN EACH OF THE FOLLOWING WORK PERIODS

	Work Periods Compared									
	1–2	1–3	1–4	1–5	1–6	1–7	1–8	1–9	1–10	1–11
Difference	−4.06	+1.02	−6.07	+1.17	+8.83	+13.80	+8.53	+5.34	+13.12	+8.14
Diff./P.E.$_{diff.}$	−0.87	+0.21	−1.33	+0.24	+1.56	+ 2.31	+1.53	+0.99	+ 2.22	+1.40

Note: The plus sign means that the coefficient of variation of a given work period is greater than that of the first work period. The minus sign means that the coefficient of variation of a given work period is less than that of the first work period.

tigued muscles, there is no evidence in this study of any systematic change in variability of output from period to period in the series of work periods. This finding suggests the question, How does the variability of work done with fatigued muscles compare with that done with unfatigued muscles when the comparison is confined to the individual work period? The computations in Table IV make this comparison possible.

TABLE IV

MEAN OUTPUT AND COEFFICIENT OF VARIATION OF WORK DONE IN EACH WORK PERIOD WITH FATIGUED MUSCLES (LAST TEN CONTRACTIONS) EXPRESSED AS A PERCENTAGE OF WORK DONE WITH UNFATIGUED MUSCLES (FIRST TEN CONTRACTIONS)

	Number of Work Period											
	1	2	3	4	5	6	7	8	9	10	11	Avg.
Mean Output (percent)	16	18	19	22	22	21	25	21	20	24	19	21
Coefficient of Variation (percent)	323	361	368	290	374	281	284	323	273	284	240	309

Note: This table is to be read as follows: The mean output of work done in the last ten contractions of the first work period, 4.68 (Table III*a*), is 16 percent of the mean output of work done in the first ten contractions of the first work period, 29.55 (Table II*a*). Likewise, the coefficient of variation of the work done in the last ten contractions of the first work period, 67.95 (Table III*a*), is 323 percent of the coefficient of variation of the work done in the first ten contractions of the first work period, 21.05 (Table II*a*).

Comparisons of variability must be limited to relative variability because of the large differences among the means. In Table IV the mean output of work done in each work period with fatigued muscles is expressed as a percentage of the mean output of work done in each work period with unfatigued muscles. Analogous comparisons are made also among the coefficients of variation. In Table IV it will be seen that, on the average, the work done in the last ten contractions of a work period in which the work is continued to exhaustion is one fifth (21 percent) of the work done in the first ten contractions. The relative variability of the work done with fatigued muscles is three times (309 percent) the variability of the work done with unfatigued muscles. An inspection of the coefficients of variation used in this comparison (Tables II*a*

and III*a*) and their respective probable errors will show that in every work period variability of the work done with fatigued muscles is significantly more variable than that done with unfatigued muscles. In finding this great variability in work performed with fatigued muscles, the present study corroborates the conclusion reached by Weinland (**4**). The great variability of work done with fatigued muscles was obscured in the first comparison, based upon total work periods (Tables I*a* and I*b*), by the fact that a rest of five minutes after exhaustion is apparently sufficient for the almost complete recovery of the flexor muscles of the middle finger.

Summary and Conclusions

The problem investigated was the effect of fatigue upon variability of output in muscular work. The data consisted of work performed by twenty-seven men college students, each of whom performed eleven periods of work or a total of two hundred ninety-seven work periods. The muscles used were the flexors of the middle finger. In each work period, work was continued to exhaustion; a rest of five minutes was interspersed between successive work periods; the load was a ten-pound weight.

The following conclusions are drawn:

1. Absolute variability (standard deviation) decreases sharply in the early work periods of a series and fluctuates irregularly throughout the remainder of the series. Relative variability (coefficient of variation) shows no regular throughout the series.

2. In work done with unfatigued muscles (first ten contractions), the work done in the later periods of the series is significantly more variable than that done in the earlier periods.

3. In work done with fatigued muscles (last ten contractions), there is no systematic change in variability throughout the series.

4. A comparison of the work done in the first ten contractions of each work period with that done in the last ten contractions shows that the mean output of work done with

fatigued muscles is one-fifth (twenty-one percent) of that done with unfatigued muscles. The relative variability of work done with fatigued muscles is three times (309 percent) the variability of the work done with unfatigued muscles.

(Manuscript received January 26, 1934)

REFERENCES

1. ASCH, S. E., An experimental study of variability in learning, *Arch. of Psychol.*, 1932, No. 143. Pp. 55.
2. MCGEOCH, J. A., Changes accompanying practice upon successive samples of verbal material, *J. Gen. Psychol.*, 1933, **9**, 117–129.
3. THORNDIKE, E. L., The Fundamentals of Learning, New York: Bureau of Publications, Teachers College, Columbia University, 1932. Pp. 638.
4. WEINLAND, J. D., Variability of performance in the curve of work, *Arch. of Psychol.*, 1927, No. 87. Pp. 68.

THRESHOLDS OF VERTICAL MOVEMENT OF THE BODY

BY HERBERT GURNEE
Western Reserve University

I. Introduction

Compared to the vast amount of data on the phenomena of rotation,[1] the amount of data on translation of the body is astonishingly small. If we exclude the work done on body-sway and the like, which as reviewed by Fearing is not at all extensive,[2] the investigations of bodily translation can be easily enumerated with one digit. Whatever the reasons for this disparity, there is no denying the fact that most of man's movement experience is essentially rectilinear, and that his primary interest is in moving from point to point rather than in circles. Much of this rectilinear movement, of course, may involve a rotary factor, but the rotary factor can hardly be significant in an arc whose center is some hundreds or thousands of feet distant.

With regard to vertical translation, there appears to be nothing of scientific import. The *Psychological Index*, from 1894 to 1931 inclusive, contains, so far as we have been able to discover, not a title that either avows or suggests an investigation of the problem. Even the work in aviation has been done on the horizontal plane, namely, in balance and in rotation. Yet the frequency, not to speak of the interest and pleasure, with which humans experience vertical movement can hardly be called negligible.

It was in order partially to fill this gap that the present investigation was begun. There were several questions that suggested themselves. At what velocity must the body be moved in order that movement may be perceived, *i.e.*, the problem of thresholds? Is there any difference in sensitivity

[1] C. R. Griffith, An historical survey of vestibular equilibration, *University of Illinois Bulletin*, Vol. XX, 1922, No. 5; *Psychol. Bull.*, 1929, **26**, 549–565.

[2] F. S. Fearing, *Jour. Comp. Psychol.*, 1924, **4**, 91–121, 163–185.

as between upward and downward movement? It seems to be common experience that downward movement in elevators, etc., is more disturbing than upward movement. "I didn't mind going up, but coming down was terrible" is a frequent remark of airplane travelers. What are the cues of vertical movement, up and down? What variability and what practice effect if any do reactions to movement stimuli show? To what extent do summation and adaptation occur? It is clear that adaptation does take place; one traveling in express elevators moving at a uniform rate very soon becomes subjectively still. Are there any after-images, such as occur in rotation? If so, what is their course? People coming ashore after several days at sea often report vivid after-images, in which the ground appears to rise and fall like the ocean. What are the reaction-times to vertical movement, up and down? Are there compensatory eye-movements, as seems most likely? If there are, what is their behavior?

Obviously these many questions were beyond the scope of a single investigation, and consequently the problem of thresholds, the relative effects of upward and downward movement, and the questions of variability and sensory cues were chosen for study.

II. Apparatus

Several kinds of apparatus were considered and finally rejected in favor of a teeter-board, in principle like the child's see-saw. This set-up of course gave movement in an arc rather than pure translation but it offered a minimum of technical difficulties and permitted a fineness of experimental control that would be probably difficult to exceed by any other device. Fig. 1 shows a schematic drawing of the arrangement. A hard pine plank, 30 cm. wide, 5 cm. thick and 630 cm. long was supported on a knife-

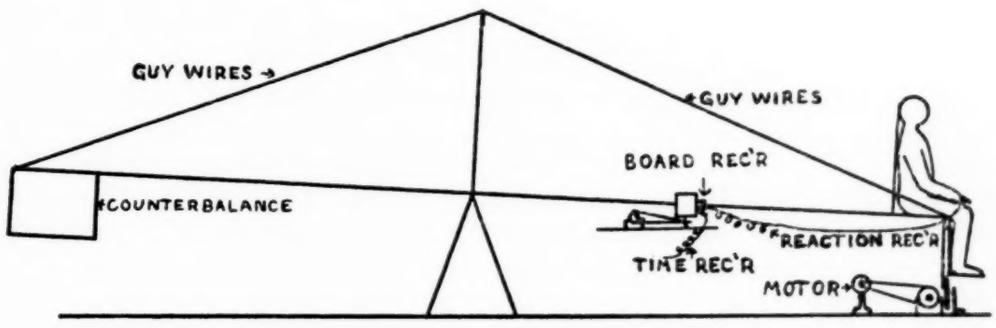


Fig. 1. Schematic drawing of apparatus set-up.

edge steel bearing, which was placed 365 cm. from one end of the plank. 52 cm. from this end, or 313 cm. from the bearing, was constructed, at right angles to the plank, a back and head rest for the subject, who sat back to the fulcrum, with legs hanging over the end of the board and supported by a foot-rest. This sitting position, which was reported comfortable, placed S's vestibular structures, as near as could be determined, 326 cm. from the fulcrum. Travis and Dodge [3] in their investigation of similar motion in the horizontal plane had found that at this distance rectilinear components predominated over rotary components. At the other end of the board was attached a large box for holding bricks with which to counterbalance the S. Bending and twisting of the plank was prevented by four guy wires, which ran from a vertical support in the center of the plank to the two ends and functioned in the manner of cantilevers. The teeter-board was moved up and down by a driving shaft, connecting S's end of the board to a driving pulley fastened to the floor. The extent of the movement could be varied by varying the distance of the driving pin from the center of the driving pulley. The driving pulley was turned by a large Westinghouse fan motor, working through a 40 to 1 reducing unit, which was provided with cone pulleys for varying the rate of movement. This set-up gave simple harmonic vertical movement, *i.e.*, movement which started from zero, accelerated to maximal velocity, then decelerated to zero again, in both upward and downward phases. This type of movement had the advantage of eliminating jars from starting and stopping the apparatus, and from changing from one movement phase to the other. It also had been used by Travis and Dodge in the study referred to. Smoked records taken of the teeter-board movement show it to be remarkably smooth. An electric fan was placed on the board just behind S to furnish a constant auditory and vibratory background, in order to mask any noises that might distract S or provide secondary cues, and to mask any vibrations that might be picked up from the building or from the motive elements. Since the building was on a very heavily used street and the apparatus was so delicately balanced it could easily be disturbed, it seemed advisable to take these precautions. The vibration produced by the fan was very faint; in fact, Ss mentioned noticing it only twice in their verbal reports. Momentum in the movement of the board was counteracted by heavy elastic bands, so adjusted that their tensions, increasing as the board decelerated, permitted the attainment of zero velocity at the end of each upward or downward movement without the disturbance of jar or recoil.

Smoked records were taken and included the board movements, the S's reactions, and a seconds time-line. A lever attached to the board gave the board movements, and a two-way electric marker gave S's reactions, which consisted in depressing two small finger keys, one for down movement, the other for up movement. This method furnished quantitative data that would have been impossible by verbal reports alone.

III. Procedure and Method

Three Ss were used, males, two of them instructors, the other a student in psychology. Both instructors had served as Ss in rotation experiments, and one, T, had been an S in the study of rotary and rectilinear oscillation by Travis and Dodge. The Ss reported at the same hour each day, namely, 11 A.M. They were blindfolded and instructed to hold the head still against the head rest and to react by depressing the right finger key if downward movement of the body was perceived, the left finger key if

[3] R. C. Travis and R. Dodge, Experimental analysis of the sensorimotor consequences of passive oscillation, rotary and rectilinear, *Psychol. Mon.*, 1928, **38**, No. 175.

upward movement was perceived, and to keep the respective key depressed as long as the movement continued. Four stimulus situations were used, in random order, three situations with movement, one without movement. The extent of movement was always 8 cm at that position, 326 cm from the fulcrum, where, so far as we could tell, the two vestibules were located. The frequencies were: a complete oscillation in 8 secs, a complete oscillation in 12 secs, and a complete oscillation in 16 secs, or a half oscillation in 4, 6, and 8 seconds respectively. These values gave average velocities of 2.0 cm per sec (16 cm in 8 sec), 1.33 cm per sec (16 cm in 12 sec), and 1.0 cm per sec (16 cm in 16 sec). At the midpoint of each half-oscillation, of course, the velocity was greater, while at the beginning and end of each half-oscillation the velocity was zero. Using the equation $V_0 = 2\pi r/T$, we find that the velocity at the midpoint was, for the 8 sec stimulus 3.14 cm per sec, for the 12 sec stimulus 2.09 cm per sec, and for the 16 sec stimulus 1.57 cm per sec. These values were selected because they promised results around the 50 percent correct point and, since three judgments were possible, no movement, upward movement, downward movement, the 50 percent point rather than the 75 percent point seemed justifiable as an arbitrary limen.

Each trial consisted of a continuous series of simple harmonic oscillations of a particular frequency, or a period of no movement, and lasted about 80 seconds. This meant 20 half-oscillations at 8 sec, 14 half-oscillations at 12 sec, and 10 half-oscillations at 16 sec per trial. Times were equated in this way in order that S might get no indication of the value of the stimulus from the duration of the trial, although as a consequence only half as many 16 sec stimuli as 8 sec stimuli could be given. In all 100 8 sec half-oscillations, 70 12 sec half-oscillations, and 50 16 sec half-oscillations were presented. About half the time a trial opened with movement on the down stroke, the remainder of the time on the up stroke, in random order. Three minutes elapsed between trials, and gave time for rest, verbal reports, and resetting of apparatus. Care was taken to make the resetting a constant feature, even before the stationary trials, when the belt was slipped and the motive elements run as usual, in spite of the masking by the fan.

The smoked records were read for the amount of perceived movement for each half-oscillation up and down, correct and incorrect, *i.e.*, right and wrong directions, and for that point in a half-oscillation where the correct movement reaction first appeared. The verbal reports gave data on cues, extent of perceived movement and other factors not obtainable from the records.

IV. Results and Discussion

Thresholds.—There are at least three ways by means of which sensitivity to vertical movement can be determined from our kind of data. Since each half-oscillation is in fact a discrete stimulus, we may find for each of the three stimulus rates the number of half-oscillations in which some movement was correctly perceived, and then convert into percentage the proportion which this number of half-oscillations bears of the total half-oscillations at that rate. There are two objections to this procedure. First, the slower stimuli because of their longer duration would afford correspondingly greater chances

of movement being perceived than the faster stimuli; and second, the amount of perceived movement would in each case be neglected, that is, a half-oscillation in which movement occurred for a fraction of a second would be given the same weight as a half-oscillation in which movement occurred for perhaps the entire phase. To partly obviate this, one might count only those half-oscillations in which movement was perceived correctly a certain proportion of the time, let us say 50 percent or more of the time, but such determinations still would not be exactly representative of the movement reactions and would be too arbitrary to be scientifically creditable. The best method seems to be to compute the amount of time movement is perceived for each stimulus-value and to convert this into percentage of the total time stimulated by that value. This has been done in Table I. The correct movement is to be found under the plus signs, the incorrect movement under the minus signs. G, for instance, when oscillated at the rate of 8 cm (16 cm in all) in 8 sec, perceives movement correctly 50 percent of the time while moving in the upward direction, 52 percent of the time while moving in the downward direction; when oscillated at the rate of 8 cm in 12 sec perceives movement correctly 26 percent of the time while moving in the upward direction, 39 percent of the time while moving in the downward direction; and when oscillated 8 cm in 16 sec, perceives movement correctly 23 percent of the time while moving in the upward direction, 36 percent of the time while moving in the downward direction. Comparison of the data for the three Ss shows the order of sensitivity, descending, to be K, T, G. It is also evident that in practically every case, though the differences are not always significant, more movement is perceived in the downward situation than in the upward situation, and it is perceived more accurately. The only important exception (the two others are insignificant) is T's direction at 16 sec. This inversion may be accidental, since the movement is too far below any acceptable threshold to be a reliable indicator of sensitivity, or it may be due to a shift in bodily cues (see below). In T's case particularly there was a notable shift from tactile to internal cues for the slow oscillations, as compared to the most rapid oscillations.

Figure 2 presents curves plotted from Table I showing the effect of decreasing the rate of movement upon the perception of movement, correct and incorrect.

If we take the point at which 50 percent correct responses occur as the threshold, we get by interpolation a rate of 8.00

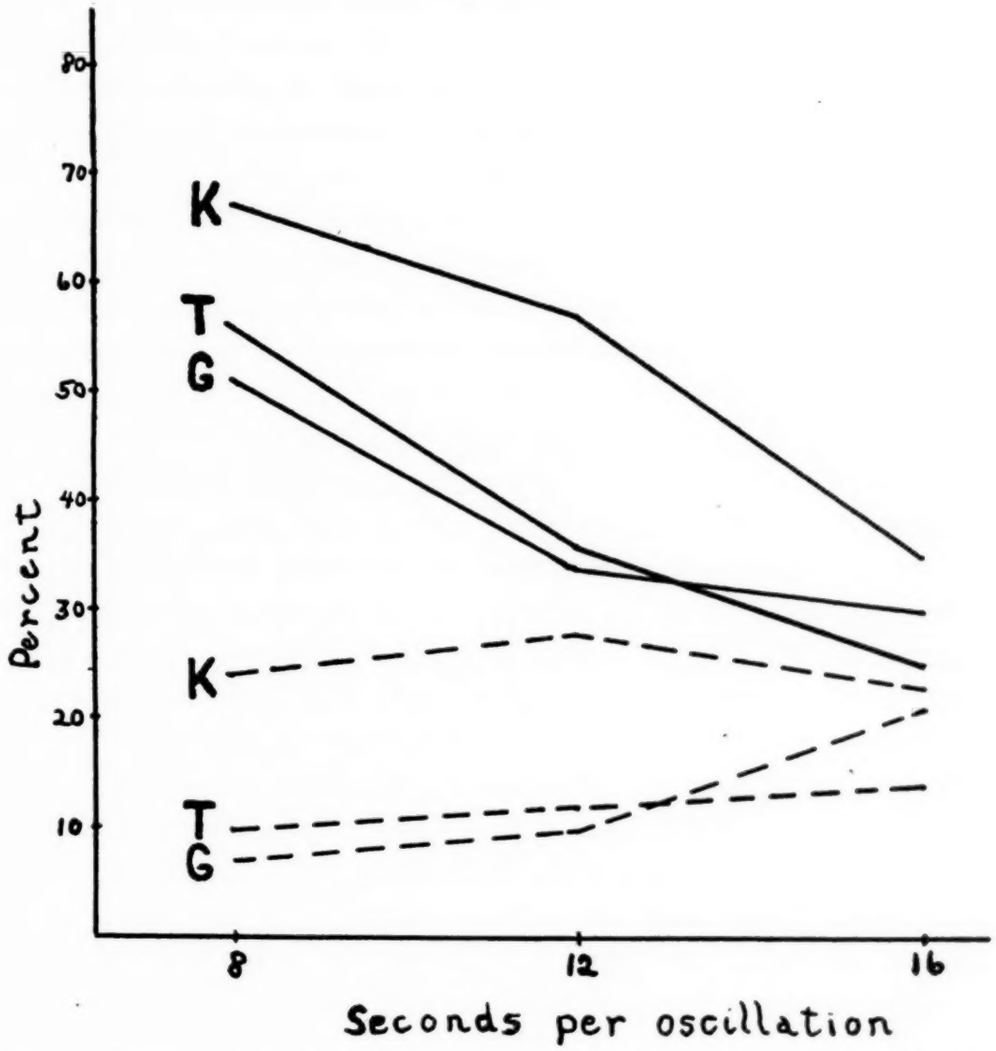


FIG. 2. Curves of perceived movement. Solid line is right direction, broken is wrong direction.

sec up, 8.61 sec down for G; 12.00 sec up, 14.26 sec down for K; and 8.27 sec up, 9.66 sec down for T. Since this means a distance of 16 cm in these times, we can calculate the average velocities, and we get, for G, 2 cm per sec for upward movement, 1.86 cm per sec for downward movement; for K, 1.33 cm per sec for upward movement, 1.12 cm per sec for down-

ward movement; and for T, 1.93 cm per sec for upward movement, 1.66 cm per sec for downward movement. If we then determine the ratios which these values give, we find, for G, sensitivity to upward movement is 93 percent of sensitivity to downward movement, for K, sensitivity to upward movement is 84 percent of sensitivity to downward movement, and for T, sensitivity to upward movement is 86 percent of sensitivity to downward movement. It is already made obvious that these ratios might not hold at all if other than the 50 percent point were taken as the threshold.

If we now examine the percentages for the negative movement, *i.e.*, movement perceived but in the wrong direction, we find an invariable tendency for the amount to increase relatively as the rate of movement decreases. We may measure this increase by determining the ratio of wrong to right movement, or by determining the ratio of wrong to total perceived movement, in each case. Adopting the latter procedure, and averaging the results, up and down for all S's, we find that wrong movement is 18 percent of the total perceived movement at 8 sec (range 11 percent to 33 percent), wrong movement is 27 percent of the total perceived movement at 12 sec (range 17 percent to 40 percent), and wrong movement is 38 percent of the total perceived movement at 16 sec (range 17 percent to 52 percent. 50 percent, of course, means a chance distribution of right and wrong directions, and this is approximated in two instances, G at 16 sec up, T at 16 sec down. It is again clear that the 16 sec stimulus is too weak to be a good measure of threshold values, at least for these Ss, though it is useful in plotting the lower end of our sensitivity curves.

These same ratios can give us a measure of the accuracy of perceived direction for each S. Averaging for all three rates, G gets 26 percent of total perceived movement wrong (range 11 to 51), T 25 percent wrong (range 12 to 52), K 33 percent wrong (range 19 to 45). It is therefore apparent that K's greater sensitivity, measured by percent of objective oscillation perceived correctly, is in part at least a function of more guessing, and it is probable that we ought to take this into

account in computing thresholds. This we might do by subtracting the wrongs from the rights for each S, plotting a curve, and finding that point at which we would have a remainder of 50 percent, or by plotting curves of wrong and right movement and finding the point at which these curves diverge by 50 percent. Following the latter procedure and computing the average velocities for the values so obtained, we get, for G 2.3 cm per sec up, 2.3 per sec down; for K 4.0 cm per sec up, 1.7 cm per sec down, for T 3.2 cm per sec up, 1.9 cm per sec down. Except that most of these values depend on extrapolation, as inspection of Fig. 2 or Table I will show, they are probably more representative of movement sensitivity than are the values based on percent of correct movement alone, if we consider direction an essential component of the movement experience.

TABLE I

PERCENTAGES OF REACTIONS TO UP AND DOWN MOVEMENT OF THE BODY, ONE COMPLETE OSCILLATION IN 8, 12, AND 16 SECONDS

Plus means perceived direction correct, minus means perceived direction incorrect.

Stimuli	G				K				T			
	Up		Down		Up		Down		Up		Down	
	+	−	+	−	+	−	+	−	+	−	+	−
8 sec	50	6	52	8	62	30	71	17	51	12	60	8
12 sec	26	11	39	8	50	34	63	21	36	13	36	11
16 sec	23	24	36	17	30	25	40	20	29	6	20	22

Variability of Responses.—The question of variability in reactions is dealt with in Table II. The data are in terms of percent of the total half-oscillations at each stimulus rate in which correct responses of the indicated durations occurred. That is, for the 8 sec stimulus G perceives correct movement of 0.1 to 1.0 sec duration in 11 percent of the half-oscillations, correct movement of 1.1 to 2.0 sec duration in 28 percent of the half-oscillations, etc. The data for up and down movements have been bunched; the differences between up and down movement were in the main the greater average duration of the down movement and the fewer downward half-oscillations in which no correct movement at all occurred. The

differences were in no case outstanding but were apparent enough to indicate superiority in favor of downward movement. The incorrect movement is not presented. The table shows a tendency for the half-oscillations in which no correct movement occurred to increase as the stimulus-intensity decreased. The O does not mean that no movement whatever occurred in the percents listed, for in many of these instances there was perceived movement in the wrong direction. The half-oscillations in which no movement at all, right or wrong, appeared were, at the 8 sec rate 5 percent, 5 percent, 3 percent, at the 12 sec rate 21 percent, 7 percent, 11 percent, at the 16 sec rate 0 percent, 31 percent, 11 percent, for G, K, and T respectively. A tendency is thus also apparent for the half-oscillations in which no perceived movement whatever occurs to increase as stimulus-intensity decreases, with one inversion at 16 sec, which, we have already seen, is an unreliable stimulus-value. K's large percentage of zero movement at 16 sec is occasioned by the fact that in one daily series, consisting of 10 half-oscillations, he failed to get a particle of movement. Reference to the verbal reports shows that movement was present but never achieved any determinable direction. In no other movement trial (single series of oscillations) did an S fail to get some directional movement.

The average duration of the movement responses obviously increased with the increase in the time of oscillation. This is to be expected. Variability in duration of responses increased similarly.

In any single half-oscillation there was almost never more than one correct response. Out of the 690 half-oscillations used in the experiment, only 6 half-oscillations are to be found, all at 16 sec, in which more than one correct response occurred, and in these 6 there were two correct responses. Evidently S reacts to a half-oscillation as a single stimulus. This fact has been observed before, in oscillation of a visual stimulus and rotary oscillation of the body and in rectilinear oscillation of the body.[4] Rarely would S change a correct to an incorrect response, and then only near the end of the

[4] Travis and Dodge, *op. cit.;* also H. Gurnee, The effect of a visual stimulus upon the perception of bodily motion, *Am. Jour. Psychol.*, 1931, 43, 26–48.

deceleration phase, often he would change an incorrect to a correct response, and this almost invariably in the acceleration phase. Sometimes a correct response would extend into the next half-oscillation and become incorrect. In fact, a considerable portion of the negative movement, especially for the more rapid rates, was due to this lag. This and uncertainty about direction at onset resulted in a preponderance of negative movement in the acceleration phase. Again, similar results have been obtained in simple harmonic oscillation of a faint visual stimulus and simple harmonic rotary oscillation of the body.

TABLE II

VARIABILITY IN MOVEMENT RESPONSES, IN TERMS OF PERCENT OF TOTAL HALF-OSCILLATIONS IN WHICH RESPONSES OCCURRED

Column at left gives duration in seconds of response.

	G			K			T		
Sec.	8	12	16	8	12	16	8	12	16
0	14	38	30	9	11	41	7	23	35
.1–1	11	7	7	2	3	5	7	14	15
1.1–2	28	23	20	15	10	5	34	22	17
2.1–3	40	15	12	38	23	5	33	20	6
3.1–4	7	9	14	36	25	13	19	14	7
4.1–5		3	9		21	8		1	17
5.1–6		5	2		7	10		6	0
6.1–7			4			3			0
7.1–8			2			10			3

Onset of Perceived Movement.—That the acceleration phase is however a more effective stimulus than the deceleration phase is clear from Table III, which shows the time after beginning of a half-oscillation at which correct movement was first perceived. Incorrect movement is not treated, and up and down data are bunched. The results as in Table 2 are in percentages of total half-oscillations for each rate. For example, taking all the 8 sec half-oscillations with which he was presented, G gets correct movement by the end of the first second in 42 percent of these half-oscillations, between 1.1 and 2 sec after onset in 28 percent of these half-oscillations, between 2.1 and 3 sec after onset in 12 percent of these half-oscillations, and between 3.1 and 4 sec after onset, or during the last

differences were in no case outstanding but were apparent enough to indicate superiority in favor of downward movement. The incorrect movement is not presented. The table shows a tendency for the half-oscillations in which no correct movement occurred to increase as the stimulus-intensity decreased. The O does not mean that no movement whatever occurred in the percents listed, for in many of these instances there was perceived movement in the wrong direction. The half-oscillations in which no movement at all, right or wrong, appeared were, at the 8 sec rate 5 percent, 5 percent, 3 percent, at the 12 sec rate 21 percent, 7 percent, 11 percent, at the 16 sec rate 0 percent, 31 percent, 11 percent, for G, K, and T respectively. A tendency is thus also apparent for the half-oscillations in which no perceived movement whatever occurs to increase as stimulus-intensity decreases, with one inversion at 16 sec, which, we have already seen, is an unreliable stimulus-value. K's large percentage of zero movement at 16 sec is occasioned by the fact that in one daily series, consisting of 10 half-oscillations, he failed to get a particle of movement. Reference to the verbal reports shows that movement was present but never achieved any determinable direction. In no other movement trial (single series of oscillations) did an S fail to get some directional movement.

The average duration of the movement responses obviously increased with the increase in the time of oscillation. This is to be expected. Variability in duration of responses increased similarly.

In any single half-oscillation there was almost never more than one correct response. Out of the 690 half-oscillations used in the experiment, only 6 half-oscillations are to be found, all at 16 sec, in which more than one correct response occurred, and in these 6 there were two correct responses. Evidently S reacts to a half-oscillation as a single stimulus. This fact has been observed before, in oscillation of a visual stimulus and rotary oscillation of the body and in rectilinear oscillation of the body.[4] Rarely would S change a correct to an incorrect response, and then only near the end of the

[4] Travis and Dodge, *op. cit.;* also H. Gurnee, The effect of a visual stimulus upon the perception of bodily motion, *Am. Jour. Psychol.*, 1931, **43**, 26–48.

deceleration phase, often he would change an incorrect to a correct response, and this almost invariably in the acceleration phase. Sometimes a correct response would extend into the next half-oscillation and become incorrect. In fact, a considerable portion of the negative movement, especially for the more rapid rates, was due to this lag. This and uncertainty about direction at onset resulted in a preponderance of negative movement in the acceleration phase. Again, similar results have been obtained in simple harmonic oscillation of a faint visual stimulus and simple harmonic rotary oscillation of the body.

TABLE II

VARIABILITY IN MOVEMENT RESPONSES, IN TERMS OF PERCENT OF TOTAL HALF-OSCILLATIONS IN WHICH RESPONSES OCCURRED

Column at left gives duration in seconds of response.

	G			K			T		
Sec	8	12	16	8	12	16	8	12	16
0	14	38	30	9	11	41	7	23	35
.1–1	11	7	7	2	3	5	7	14	15
1.1–2	28	23	20	15	10	5	34	22	17
2.1–3	40	15	12	38	23	5	33	20	6
3.1–4	7	9	14	36	25	13	19	14	7
4.1–5		3	9		21	8		1	17
5.1–6		5	2		7	10		6	0
6.1–7			4			3			0
7.1–8			2			10			3

Onset of Perceived Movement.—That the acceleration phase is however a more effective stimulus than the deceleration phase is clear from Table III, which shows the time after beginning of a half-oscillation at which correct movement was first perceived. Incorrect movement is not treated, and up and down data are bunched. The results as in Table 2 are in percentages of total half-oscillations for each rate. For example, taking all the 8 sec half-oscillations with which he was presented, G gets correct movement by the end of the first second in 42 percent of these half-oscillations, between 1.1 and 2 sec after onset in 28 percent of these half-oscillations, between 2.1 and 3 sec after onset in 12 percent of these half-oscillations, and between 3.1 and 4 sec after onset, or during the last

second of half-oscillation in 4 percent of these half-oscillations. The zero values at the bottom of the table are the same as those in the top row of Table II and mean, of course, that a correct response did not begin at any time in the given percent of half-oscillations.

It is evident that correct responses usually began in the acceleration phase, the end of the acceleration phase coming at 2 sec for the 8 sec stimulus, at 3 sec for the 12 sec stimulus, and at 4 sec for the 16 sec stimulus. For G, K, and T respectively we have the following percents of half-oscillations in which correct responses began in the acceleration phase: 70 percent, 76 percent, 74 percent at 8 sec, 54 percent, 70 percent, 56 percent at 12 sec, 44 percent, 37 percent, 43 percent at 16 sec, and in the deceleration phase 16 percent, 15 percent, 19 percent at 8 sec, 8 percent, 19 percent, 21 percent at 12 sec, 26 percent, 22 percent, 22 percent at 16 sec. It is obvious that the advantage of the acceleration phase decreases as the rate of acceleration and deceleration decreases. The average time in seconds at which correct movement first appeared is given in the bottom row of the table. The time is seen to be later as the rate decreases, though it always remains in the acceleration phase. Averaging the results of the 3 Ss, we get 1.23 sec for the 8 sec stimulus, 1.87 sec for the 12 sec stimulus, and 3.20 sec for the 16 sec stimulus. Using the equation $V = \frac{Y}{r} V_0$ for determining the velocity at different points in simple harmonic motion,[5] we learn that the velocities attained at these particular points are 2.55 cm per sec for the 8 sec stimulus, 1.75 cm per sec for the 12 sec stimulus, and 1.49 cm per sec for the 16 sec stimulus. Evidently as the rate of movement decreases the attained velocity necessary to induce a movement response also decreases. This apparent paradox can probably be explained by the principle of summation, since, we observe, the slower the rate the longer the time necessary to reach the critical point, the 1.49 cm per sec velocity in the case of the 16 sec stimulus

[5] $V = V_0 \sin e$ will give the same results. See A. L. Kimball, *College Physics*, 1929.

requiring almost 3 times as long to attain as the 2.55 cm per sec velocity in the case of the 8 sec stimulus.

These same values might be used as a measure of S's sensitivity except that they do not take into account the half-oscillations in which no correct movement occurred.

TABLE III

TIME AFTER ONSET OF HALF-OSCILLATION AT WHICH CORRECT MOVEMENT WAS FIRST PERCEIVED

Given in terms of percent of total number of half-oscillations.

	G			K			T		
Sec.	8	12	16	8	12	16	8	12	16
.1–1	42	27	24	44	21	11	51	29	12
1.1–2	28	16	6	32	32	8	23	18	5
2.1–3	12	11	8	13	17	6	17	9	16
3.1–4	4	8	6	2	14	12	2	6	10
4.1–5		0	6		4	9		12	8
5.1–6		0	17		1	8		3	7
6.1–7			3			3			5
7.1–8			0			2			2
0	14	38	30	9	11	41	7	23	35
Av.	1.3	1.6	2.9	1.2	2.0	3.3	1.2	2.0	3.4

There is an indeterminate error in the results in Table III, namely, the latency of the recording mechanism, and of S's reactions. The latency of the former is probably negligible and would not make a perceptible difference in the records. Respecting the latter there seems to be no satisfactory answer. We can only say that our data give the times of S's manual reactions and we must leave the question of when S may have first perceived movement to the realm of speculation.

Extent of Movement.—Judgments of extent of perceived movement varied from less than an inch to several feet. Infrequently S would be about 'to hit the ceiling' or 'strike the floor.' The greater amplitudes came during a prolonged response that sometimes covered two or more half-oscillations and were increasingly frequent as the rate decreased. As a rule judgments ranged between 3 and 10 inches, there being a noticeable tendency to overestimate rather than underestimate the amount of objective displacement.

Sensory Cues.—Movement cues were reported as pressure and kinæsthesis. At 8 sec cutaneous pressures predominated, generally from the seat, with increased pressure going up, decreased pressure going down. As the stimulus rate decreased, internal cues became more and more evident. These were almost invariably a vague, diffuse, general kinæsthesis and a diffuse pressury sensation from within the head. The general kinæsthesis was more frequently mentioned. It seemed usually to be 'all over'; "Wherever I look, I can feel it." Sometimes the cues were too vague and faint to be recognizable; "I don't know, I just seemed to be moving." Movement without direction often occurred and seemed to be mediated by a 'light, pressury liveliness.' Visceral cues were rarely reported. This is surprising in view of the frequency and intensity with which they apparently occur in moving elevators, swings, etc. The difference is probably a function of different stimulus intensity, our movement being evidently too slow to excite visceral disturbances. Kinæsthesis of eye-movements was never mentioned in the reports. In rotary oscillation on the other hand sensations of eye-movements are frequently mentioned. Under what conditions compensatory eye-movements may occur during vertical movement of the body is not yet known. It may be that our stimuli were inadequate or that ocular movements occurred but were too slight to be perceptible. The problem is one for investigation with photographic technique.

Breathing occasionally interfered with the movement experience. Shallow breathing had little effect but deeper breathing resulted in noticeable facilitation when the direction of chest movement coincided with the direction of body movement and noticeable inhibition when the direction of chest movement was contrary to body movement. That is, inhalation facilitated, exhalation inhibited upward movement, and vice versa.

It appears unwise to draw any conclusions regarding the probable role of the vestibule in vertical movement experiences. It may be that the head pressures reported are the result of vestibular stimulation and it may be that the vague

general kinæsthesis so commonly referred to is a reflex consequence of vestibular stimulation, but we prefer to suspend judgments until more direct evidence is in.

Hallucinatory Movement.—In the still trials, one of which was given in each experimental session, there was usually some perceived movement, varying from individual to individual and from day to day. Much of this movement was without direction and therefore not recorded in the smoked records. For the directional movement we have the following data. G moved 38 percent of the time, 15 percent up and 23 percent down, K moved 18 percent of the time, 6 percent up and 12 percent down, and T moved 32 percent of the time, 19 percent up and 13 percent down. The down obviously has the advantage for G and K, as in the movement trials, the up for T, as in the 16 sec trials. The frequency of the responses varied from trial to trial, G getting from 6 to 20 in a single trial, T from 3 to 8, and K from 0 to 5. K was the only S to go through a complete still trial without getting any directional movement, and he did this twice. G and T because of previous experimental work possessed movement sophistication, and this may have had something to do with their results.

The hallucinatory responses were generally oscillatory. The durations of single responses ranged from 1 sec to 14 sec for G, median 2 sec, 2 sec to 15 sec for K, median 6 sec, and 1 sec to 9 sec for T, median 3 sec. The cues were invariably, where reportable, a vague general kinæsthesis, as in the 16 sec trials.

It is interesting to note that no directional movement other than up and down was reported for these still trials. In studies of horizontal movement on the other hand, especially in studies of rotary oscillation, the still trials have yielded reports of only horizontal movement. This looks clearly like the influence of set. Whatever the origin of the obscure kinæsthetic factors, they are evidently interpreted as movement sensations and given direction in accordance with the prevailing attitude, up and down, right and left, to and fro, as the case may be.

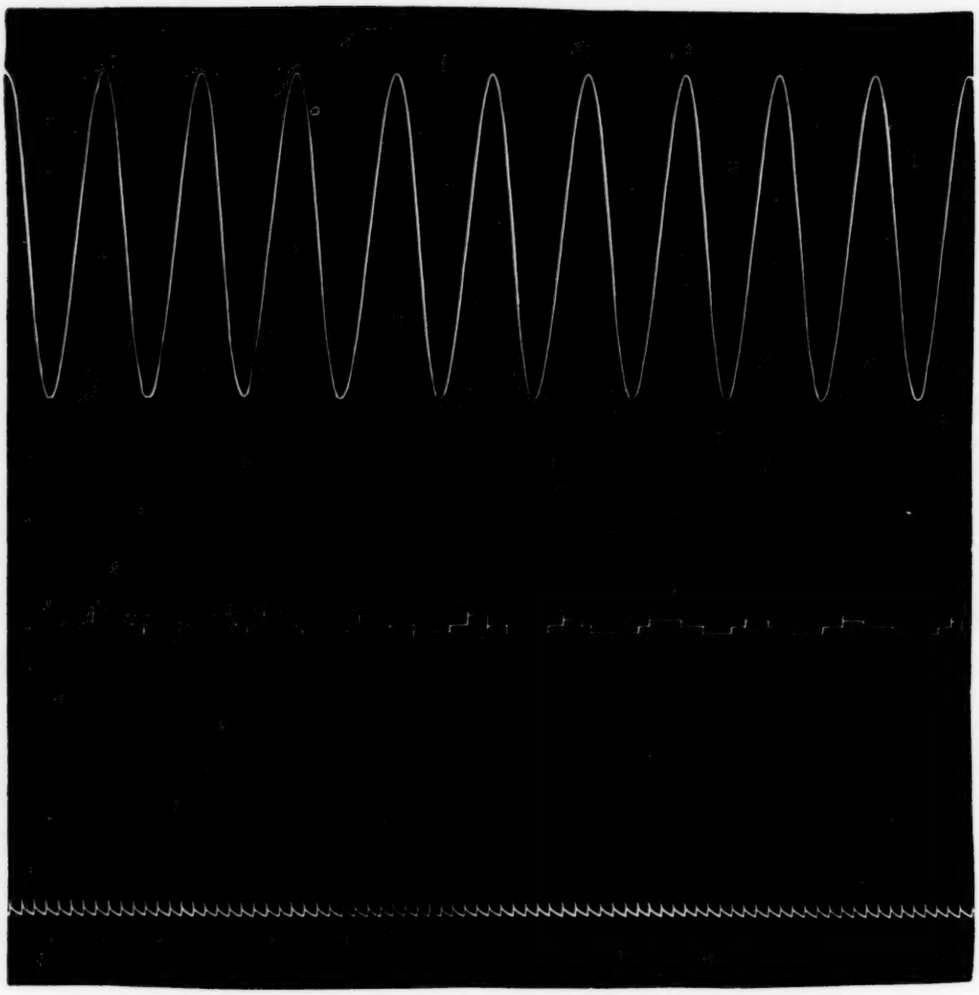


FIG. 3. A typical record. The curves represent movement of the teeter-board. The time line is in seconds.

V. SUMMARY

1. Simple harmonic motion was used to oscillate vertically 3 Ss 8 cm in 8 sec, in 12 sec, and in 16 sec.

2. Thresholds, chosen as the point of 50 percent correct responses, ranged for the three Ss between 1.12 cm per sec and 2.0 cm per sec, average velocity. Subtracting wrong movement from right movement, we obtained thresholds ranging between 1.7 cm per sec and 4.0 cm per sec., average velocity.

3. Sensitivity to upward movement was somewhat less than sensitivity to downward movement, the difference ranging between 84 percent and 93 percent for the 3 Ss.

4. Percent of right responses increased, percent of wrong responses (movement perceived but in the wrong direction) decreased with increased velocity.

5. Ss varied from oscillation to oscillation and day to day in duration, estimated extent, and accuracy of movement. Variability decreased as velocity increased.

6. Correct responses usually began in the positive acceleration phase. There was evidence of summation in the positive acceleration phase.

7. Usual cues were cutaneous pressure and a general kinæsthesis, with cutaneous pressures predominating at the fastest rate, general kinæsthesis at the slowest rate.

8. Hallucinatory movement occurred in the still periods, from 18 percent to 38 percent of the time for the 3 Ss. The explanation proposed for this hallucinatory movement was interpretation of some vague kinæsthetic factor in terms of the mental set.

(Manuscript received February 16, 1933)

JUDGMENT TIMES OF DIFFERENT PSYCHOPHYSICAL CATEGORIES

BY S. W. FERNBERGER, E. GLASS, I. HOFFMAN AND M. WILLIG

University of Pennsylvania

Divergent results have been obtained for judgment times of the different categories of judgment in psychophysical experiments performed by different experimenters. Martin and Müller,[1] George [2] and Kellogg [3] all found longer times for the equal judgments than for those either of less or greater. Fernberger and Irwin [4] for the method of single stimuli and Carlson, Driver and Preston [5] for the method of constant stimuli obtained judgment times for all three categories which were exceedingly similar. In discussing these differences the latter authors suggest the possibility that the differences between their results and those of other investigators may be due to differences in the attitude of the observers toward the equality judgments. They say: "In both George's and Kellogg's experiments, it would seem that there was set up an attitude against giving equality judgments. This would be accomplished in the case of George by adding a large number of 'doubtful' and 'guess' categories and, in the case of Kellogg, by introducing a 'two-category' series in which equal judgments were eliminated by experimental direction. All observers in the present experiment had been previously trained in an attitude in which the equal judgment was accepted as

[1] L. J. Martin and G. E. Müller, *Zur Analyse der Unterschiedsempfindlichkeit*, 1899, 1–96.

[2] S. S. George, Attitude in relation to the psychophysical judgment. *Amer. J. Psychol.*, 1917, **28**, 1–37.

[3] W. N. Kellogg, The time of judgment in psychometric measures. *Amer. J. Psychol.*, 1931, **43**, 65–86.

[4] S. W. Fernberger and F. W. Irwin, Time relations for the different categories of judgment in the 'absolute method' in psychophysics. *Amer. J. Psychol.*, 1932, **44**, 505–525.

[5] W. R. Carlson, R. C. Driver and M. G. Preston, Judgment times for the method of constant stimuli. This *Journal*, 1934, **17**, 113–118.

a category equally valid and quite on a par with either of the categories of difference judgments" (pp. 117).

In order to test out this hypothesis the present experiment was planned. Fernberger [6] has shown that the attitude of the observer toward the equality judgment has a profound influence on the relative number of such judgments which will be given in an investigation and he has also demonstrated that these differences in attitude may be effected by the instructions given to the observers. It was planned to determine if changed instructions which would create an attitude against giving equal judgments might have an effect also on the relative judgment times for the three categories.

The observers in the present experiment were all naive and totally without practice in the giving of psychophysical judgments. The following instructions were read by the observers and discussed by the senior experimenter until he was sure that the last sentence especially had been impressed upon them.

"You will be presented with weights in pairs—first a standard weight and then a comparison weight. These pairs will be judged separately, that is, each comparison weight will be judged in terms of its own standard weight. In some cases, the second weight will seem heavier than its standard and in this case you will give the judgment 'heavier'; in some cases the second weight will seem lighter than its standard and in this case you will give the judgment 'lighter.' In some few cases, the second weight may seem equal to the standard and in other cases it may be that you will be unable to determine a difference between the two weights. If the two weights of any pair should seem actually equal or if you cannot determine a difference between the two you will report 'equal.' *But a judgment of lighter or heavier should be given if at all possible.*"

At the beginning of the experiment, we employed exactly the same experimental arrangement used by Carlson, Driver and Preston except for the differences in instructions. The weight stimuli were arranged about the circumference of a turning top table which was wired with a Dunlap chronoscope so that the lifting of each comparison weight broke a circuit and actuated the chronoscope. The observer pressed a bulb

[6] S. W. Fernberger, The effect of the attitude of the subject upon the measure of sensitivity. *Amer. J. Psychol.*, 1914, **25**, 538–543; The use of equality judgments in psychophysical procedures. *Psychol. Rev.*, 1930, **37**, 107–112; Instructions and the psychophysical limen. *Amer. J. Psychol.*, 1931, **43**, 361–376; Die Ungewissheitsurteile in der Psychophysik. *Arch. f. d. ges. Psychol.*, 1931, **80**, 273–290.

key to stop the chronoscope. Readings were taken in units of 1/1200 second. With the end of the third series, it had become evident, and this was confirmed by the introspections of the observers, that they were anticipating the judgment by pressing the bulb too soon and before the judgment had been completely formulated. To avoid this error a change was made in the experimental arrangement. For the last five series for each observer a voice key was substituted for the bulb key. In this way both the actuation and stopping of the chronoscope was made completely automatic—the lifting of the comparison weight started the chronoscope and the actual verbalization of the judgment into the voice key stopped the chronoscope.

All of the three junior authors acted both as experimenters and as observers. At the beginning of the experiment they were totally unacquainted with psychophysical techniques and with the formulation of judgments on small differences in weight. All observers worked without knowledge of the results and of the problem until the end of the experiments. All three observers were undergraduate seniors majoring in psychology at the University of Pennsylvania. From each observer were obtained 8 series, each representing 350 judgments. So this experiment is based upon 2,800 judgment times for each observer or a total of 7,400 judgment times for the entire experiment. The material is more extensive than that of Martin and Müller, George or Kellogg.

The results will be found in Tables 1–3, one table being given to the data for each observer. Each table contains the number of judgments, the average judgment times and the sigma of the average of these reaction times for each of the three categories of judgment and for every series. The series taken with the bulb key (Series 1–3) and those obtained with the use of the voice key (Series 4–8) are separated. It will be noted that the number of equal judgments is smaller than one usually obtains in such an experiment. This is evidence of the effectiveness of our instructions in creating an attitude against the giving of judgments of this category. It will also be observed, in the case of all three observers, that the judg-

TABLE 1

Ob. G. Reaction Times

Series	Lighter			Equal			Heavier		
	N	Ave	σ_{ave}	*N*	Ave	σ_{ave}	*N*	Ave	σ_{ave}
1	142	832	57.90	29	1174	150.30	179	763	49.30
2	130	217	7.45	35	202	11.47	185	229	6.70
3	150	268	5.96	27	251	8.49	173	262	5.44
4	156	467	8.26	22	555	28.80	172	450	7.31
5	173	532	8.80	33	626	35.17	144	538	11.85
6	173	492	6.83	20	563	43.91	157	462	8.12
7	172	535	5.61	19	591	28.49	159	531	8.11
8	151	562	7.21	21	584	34.41	178	524	4.93

TABLE 2

Ob. H. Reaction Times

Series	Lighter			Equal			Heavier		
	N	Ave	σ_{ave}	*N*	Ave	σ_{ave}	*N*	Ave	σ_{ave}
1	175	673	45.93	52	914	96.39	123	690	62.83
2	185	244	7.64	39	211	10.57	126	247	8.22
3	130	260	7.56	22	254	19.47	198	231	5.06
4	200	394	7.31	18	503	44.71	132	374	8.45
5	223	482	6.81	17	520	35.70	110	482	10.14
6	192	500	7.35	29	453	18.21	129	487	9.96
7	174	536	6.56	15	510	28.35	160	519	6.20
8	156	655	10.52	16	517	43.82	178	620	8.61

TABLE 3

Ob. W. Reaction Times

Series	Lighter			Equal			Heavier		
	N	Ave	σ_{ave}	*N*	Ave	σ_{ave}	*N*	Ave	σ_{ave}
1	144	441	31.79	64	372	12.47	142	402	27.79
2	136	379	6.50	50	422	14.35	164	380	6.86
3	171	350	3.94	36	376	10.24	143	350	4.29
4	150	471	6.59	40	469	18.26	160	447	5.39
5	154	541	8.24	30	598	26.94	166	492	6.27
6	180	464	7.32	38	483	20.24	132	475	7.59
7	168	512	7.72	33	582	27.35	149	539	9.85
8	163	556	10.20	40	662	22.86	147	537	8.82

ment times become longer with the abandonment of the bulb key and the substitution of the voice key.

An examination of these tables at once indicates certain rather striking results. At the beginning of experimentation, in Series 1, the judgment times for the equal judgments were strikingly longer for Obs. G and H but not for Ob. W. Also it will be seen that Ob. W. gave more equal judgments than either of the other two observers. The senior author, who believed that Ob. W. was not acting under the instructions, further emphasized the last sentence with the effect that, in Series 2 for Ob. W., these time relations were reversed. It will also be noticed that the sigma of the average for the equal judgments is always materially larger than the similar values for either of the difference judgments. Analysis shows that this is due, not to greater variability of judgment times but to the fact that a smaller number of judgments entered into the calculations of these values.

A comparison of reaction times for the different categories of judgments can best be made in terms of a consideration of the differences between the averages and of the values of P_z for each of the paired differences. The values of P_z are found by dividing the difference by the probable error of the difference and then with the use of the table in Johnson.[7] The value of P_z is an index of the probability that the difference is not due to chance.

The values of both the differences and of P_z are found in Tables 4–6. In every case, in the first two pairs of columns (Lighter-Equal or Equal-Heavier) the results in bold type indicate that the difference judgment time was longer than that for the equal judgment. In all cases of normal type, the equal judgment times were longer than those for the difference judgment.

The results for Ob. G. (Table 4) are relatively clear cut. For Series 1 the equal judgment times are very considerably longer than those either for lighter or heavier. In the next two series, the last with the bulb key, the equal judgment times are smaller in both cases. For the last five series, with

[7] W. W. Johnson, *The Theory of Errors and Method of Least Squares.* N. Y.: 1892. Pp. 154.

TABLE 4

OB. G.

Series	Lighter-Equal		Equal-Heavier		Lighter-Heavier	
	D	P_z	D	P_z	D	P_z
1	342	0.8492	411	0.9217	69	0.4606
2	**15**	0.5460	**27**	0.8291	12	0.5817
3	**17**	0.7313	**11**	0.5378	6	0.3823
4	88	0.9534	105	0.9822	17	0.7011
5	94	0.9194	88	0.8901	6	0.2179
6	71	0.7195	101	0.8726	30	0.9314
7	56	0.8070	60	0.8291	4	0.2179
8	22	0.3270	60	0.7567	38	0.9968

TABLE 5

OB. H.

Series	Lighter-Equal		Equal-Heavier		Lighter-Heavier	
	D	P_z	D	P_z	D	P_z
1	241	0.8726	224	0.8116	17	0.1180
2	**33**	0.9121	**36**	0.9304	3	0.1445
3	**6**	0.1551	23	0.5581	29	0.9686
4	109	0.8959	129	0.9446	20	0.7727
5	38	0.5212	38	0.5085	0	0.0000
6	**47**	0.8930	**34**	0.7313	13	0.5212
7	**26**	0.4517	**9**	0.1656	17	0.7952
8	138	0.9610	103	0.9121	35	0.9176

TABLE 6

OB. W.

Series	Lighter-Equal		Equal-Heavier		Lighter-Heavier	
	D	P_z	D	P_z	D	P_z
1	**69**	0.8270	30	0.4914	39	0.4651
2	43	0.9344	42	0.9250	1	0.0591
3	26	0.8901	26	0.8855	0	0.0000
4	**2**	0.0538	22	0.5660	24	0.9428
5	57	0.8270	106	0.9902	49	0.9986
6	19	0.4472	8	0.1971	11	0.5170
7	70	0.9029	43	0.6818	37	0.9541
8	106	0.9958	125	0.9994	19	0.6584

the voice key, the equal judgment times are again invariably longer than those of either of the difference judgments. Except for the first series, the judgment times for the heavier-lighter categories are very similar. Although, for this observer, the values of P_z never reach unity, this representing with absolute certainty that these differences are not due to chance, still practically all of the values are exceedingly high and, in all but two cases, are higher than the corresponding values of P_z for the lighter-heavier difference. It will be remembered that the values of the sigma of the average for the equal judgments are relatively large due to the small number of cases and, therefore, it seems rather significant that the values of P_z for the lighter-equal and for the equal-heavier differences should be so nearly uniformly larger than those for the lighter-heavier differences where the values of the sigma of the average were so much smaller in both cases.

These same values for Ob. H. (Table 5) indicate certain points of agreement and also of disagreement with those of Ob. G. Like the former results, the differences between the judgment times for the lighter-heavier comparison are very small and, on the whole, the values of P_z for these differences are small. For Series 1 the judgment times for the equal category are considerably larger than those for either of the difference categories and the values of P_z are relatively high. For Series 2 and 3 (the last two series with the bulb key) the lighter judgment times are longer than the equal in both cases and the heavier times are longer than the equal in Series 2 and not very different in Series 3. In the first two series with the voice key (Series 4 and 5) the times for the equal judgments are again longer than those for either of the two difference judgments—very much longer in the case of Series 4 where again the values of P_z are high. In the last three series (Series 6–8) the equal times are invariably shorter than those for either of the difference judgments and this difference is very marked in the last series. It will be remembered that this observer gave the smallest number of equal judgments of any of the three subjects.

In the case of Ob. W. (Table 6) a somewhat similar situa-

tion is evident. It was this observer who was not living up to the instructions against giving equal judgments in the first series to the same extent, at least, as the other two observers. In this first series the times for the equal judgments are smaller than those for the lighter judgments but longer than those for the heavier. In every other case but one (lighter-equal, Series 4) the equal judgments are longer than those for either of the other two difference categories. Great variability is found in the size of the values of P_z but 9 out of 12 of these values are larger than 0.82. On the whole, and with only four exceptions out of 14 cases, the differences between the times of the lighter-heavier comparisons are smaller than those involving the equal judgment. Again great variability is found in the magnitude of the values of P_z in the cases of the lighter-heavier comparison.

Conclusion.—It has been found that, with the use of an experimental arrangement such as that employed by Fernberger and Irwin and by Carlson, Driver and Preston, who both obtained no difference between the judgment times for the three categories of psychophysical judgment, that the judgment times for the equal judgment become significantly longer, on the whole, than judgments of either less or greater if the attitude is created in the observers of not giving equal judgments unless absolutely necessary. This attitude can be created by the use of appropriate instructions. It would seem that this experiment adequately explains the differences in the results of judgment times found by Fernberger and Irwin and by Carlson, Driver and Preston, on the one hand, and by Martin and Müller, George and by Kellogg, on the other.

(Manuscript received January 25, 1934)

THE RELIABILITY OF NONSENSE-SYLLABLE SCORES

BY J. B. STROUD, A. F. LEHMAN, AND CELIA McCUE

Kansas State Teachers College

This experiment is devoted to the study of the effect of (1) length of list, (2) practice, and (3) degree of learning, upon the reliability of nonsense syllable scores. The writers also hope that this experiment will, incidentally, throw further light upon the general question of the reliability of nonsense-syllable scores.

Length of list, practice, and degree of learning are variable factors in experiments involving the use of the nonsense syllable. Different experimenters use lists of different lengths. Varying degrees of practice have been required of subjects as a prerequisite to participation in psychological experiments. Different standards of learning have been adhered to. Some experimenters have required their subjects to learn the material to the point of one correct repetition. Others have required two and three correct repetitions. It seems desirable therefore, to know the effect of these variations in experimental procedure upon the reliability of the scores obtained. A knowledge of any factors which may influence the reliability of the nonsense syllable will add to the merit of experiments involving its use.

Procedure

It was possible so to arrange the experimental procedure as to make a simultaneous investigation of the foregoing problems. The subjects were required to learn six lists of syllables—one list a day for six consecutive days. There were two lists each of 6, 12, and 18 syllables. Hence it was possible to determine the effect of length of list upon reliability by correlating paired scores for comparable lists. The effect of practice upon reliability was determined by

correlating the scores of the first day with those of the second day, those of the third day with the fourth and those of the fifth day with the sixth. Furthermore, this procedure enabled the writers to compute reliability coefficients for each of the six days by correlating the number of trials required to learn the odd-numbered items with that required to learn the even-numbered items. The syllables were learned to the point of three consecutive errorless repetitions. By keeping complete experimental records it was possible to correlate the scores of comparable lists learned to this or to any lesser degree. Such correlations were computed for the following degrees of learning, right but two, right but one, all right, and all right twice.

Three-letter syllables were used throughout. They were presented by the method of anticipation. The subjects spelled them out upon presentation and recalled them verbally. The syllables were presented at a rate of about one per three seconds. A maximum interval of approximately six seconds was allowed for recall. Since a certain amount of variability was allowed in rate of recall, the results are treated in terms of both time and trial scores. Time and place conditions were kept constant. Seventy-six upperclassmen and graduate students comprised the subjects.

Twenty-five subjects learned the lists in the order, 6, 6, 12, 12, 18, 18; 25 learned them in the order 12, 12, 18, 18, 6, 6; and 26 learned them in the order 18, 18, 6, 6, 12, 12. The 72 syllables used in the experiment were shifted systematically throughout the lists for the purpose of counterbalancing difficulty. Since the comparative difficulty of the lists was unknown it seemed imperative to do this in order to prevent the possibility of differences in results being due to differences in similarity of the respective paired lists.

Results and Discussion

(1) *Effect of Length of List upon the Reliability of Nonsense-Syllable Scores*

Two methods were employed in studying the relation of length of list to the reliability of the scores. The first is the

method of correlation; the second is the method of comparing the coefficients of variability and the PE's of the various lists. Table I shows the correlation between the scores of the comparable lists for both time and trials.

TABLE I

CORRELATION BETWEEN THE SCORES OF PAIRED LISTS OF NONSENSE SYLLABLES FOR EACH LENGTH OF LIST

Variables	Scores			
	Trials		Time	
	r	P.E.	*r*	P.E.
6 with 6	.61	.048	.68	.041
12 with 12	.73	.036	.87	.018
18 with 18	.77	.031	.62	.048

With one notable exception these results point to greater reliability of the longer lists. It is a significant fact that in the case of both time and trial scores the coefficients for the 12-syllable lists are markedly greater than those of the 6-syllable lists. It is not clear whether the reliability of the time scores diminishes at the greater lengths, as $r = .62$ would indicate, or whether this low correlation is due to chance. The other measures of reliability used in this study show no falling off in reliability of time scores of the 18-syllable lists.

While it seems that the method of correlating the number of trials required to learn the even-numbered items with that required to learn the odd-numbered items will not give a true measure of reliability, it does appear to be serviceable for comparative purposes. The main function of reliability coefficients in this case seems to be that of their predictive significance respecting subsequent performance on a comparable list of syllables. Thus correlation between two comparable lists of syllables is regarded as the truest measure of reliability.

Since each subject learned two lists of each 6, 12, and 18 syllables, 152 cases are represented in each of the correlations between odd and even-numbered items. Table II gives the correlations for each of the three lengths of lists used.

TABLE II

CORRELATIONS BETWEEN NUMBER OF TRIALS REQUIRED TO LEARN ODD- AND EVEN-NUMBERED SYLLABLES

Variables	*r*	P.E.	*n*
Odd *vs.* even 6's	.65	.015	152
Odd *vs.* even 12's	.95	.005	152
Odd *vs.* even 18's	.96	.005	152

These correlations indicate greater reliability of trial scores for the 12 and 18-syllable lists than that for the 6-syllable lists.

The great difference in magnitude of the coefficients between comparable lists and those between odd- and even-numbered items presents an interesting psychological problem. Owing to the system of shifting syllables employed it can scarcely be said that a greater similarity existed between the odd- and even-numbered items than between the items of two different lists. The interpretation must be sought in terms of the psychological processes employed. Variation in the mental and physiological condition of the subjects which occurs from one experimental sitting to another is practically ruled out where the odd- and even-numbered items are considered. It would seem to be an important question as to whether lack of a perfect correlation between the scores of two lists of nonsense syllables of the same length is due to incomparability of the two lists or to individual variation from one sitting to another. It appears that if individual variation lowers the reliability of the scores there ought to be a higher correlation between the trials required to learn the first and second 6 syllables of a 12-syllable list, for example, than between two lists of 6 syllables each learned at a different time. Such a correlation was computed using both lists of 12 syllables, making 152 cases. A coefficient of .65 $\pm$ .031 was obtained. This can scarcely be said to be significantly greater than the correlation of .61 obtained between the trials required to learn the two lists of 6 syllables. It is the opinion of the writers, however, that it would have been much greater had it not been for wide variation in 'order of learning.' While there was a strong tendency to favor the initial syllables as shown by the fact that an average of 5.6 trials were required

to learn the first six syllables (of the 12 syllable lists) and 8.2 trials to learn the second six, there were pronounced individual variations in 'order of learning.' This undoubtedly lowered the correlation.

Finally the relation between length of list and reliability is expressed in terms of coefficients of variation. The results are shown in the columns σ 100/M, Table III. Means and P.E.'s are also given.

TABLE III

MEANS, P.E.'s, SIGMAS, AND COEFFICIENTS OF VARIABILITY FOR EACH LIST

List	Trials				Time			
	M	P.E.	σ	σ 100/M	M	P.E.	σ	σ 100/M
(1) 6.....	5.90	.15	1.97	33.39	3.25	.17	2.17	66.77
(2) 6.....	5.60	.16	2.06	36.78	2.80	.13	1.75	62.50
(1) 12.....	10.50	.28	3.74	35.62	18.24	.75	9.84	53.95
(2) 12.....	10.11	.27	3.56	35.21	16.17	.69	9.03	55.84
(1) 18.....	14.92	.30	3.95	26.47	32.25	.78	10.20	31.60
(2) 18.....	13.79	.33	4.27	30.96	26.91	.81	10.58	39.32

These results generally show greater reliability for the longer lists. There is one exception. This measure fails to show greater reliability of trial scores of the 12-syllable lists than that for the 6-syllable lists. On the basis of trial scores these data show (1) 12 to be .72 as reliable as (1) 18. Similarly they show (2) 12 to be .78 as reliable as (2) 18. Likewise, (1) 6 is .79 as reliable as (1) 18, while (2) 6 is .72 as reliable as (2) 18. Upon the basis of similar calculations for time scores (1) 6 is .81 as reliable as (1) 12. (2) 6 is .88 as reliable as (2) 12. (1) 12 is .59 as reliable as (1) 18. (2) 12 is .72 as reliable as (2) 18.

Attention is called to the fact that the P.E.'s of the 18-syllable lists are very little larger than those of the 12-syllable lists. This, of course, indicates greater reliability of the 18-syllable lists.

The various techniques utilized in studying the relation of reliability to length of list show, in the main, that lists of 12 and 18 syllables yield more reliable scores than lists of 6 syllables. The data presented in Table III show less relative

variability in scores of 18-syllable list than in those of the 12-syllable lists. The correlations between comparable lists, Table I, fail to show any advantage of the 18-syllable lists over the 12-syllable lists when time scores are considered. They show a very small advantage for the 18-syllable lists over the 12-syllable lists when trials are considered.

The correlations, as a whole, and the rather close correspondence between the means of comparable lists, especially when practice effect on the second list of each pair is discounted indicate more or less satisfactory reliability of nonsense syllable scores.

(2) *The Effect of Practice upon the Reliability of Nonsense-Syllable Scores*

The first method of studying this problem consists of correlating the scores for the first day with those of the second day, the scores of the third day with those of the fourth and those of the fifth day with the scores of the sixth. The amount of practice represented in the first four days is as much as that usually required preliminary to participation in psychological experiments. Its influence ought to be apparent in the correlations between the scores of the fifth and sixth days, if it has any effect upon reliability. Each correlation given in Table IV is based upon paired scores of the three lengths of lists used, *i.e.* differences in length of list are ignored This, of course, is conducive to abnormally high correlations. However, since this spurious factor is constant it is not a limitation to the procedure herein employed. Seventy-six cases are represented in each of the correlations.

TABLE IV

CORRELATIONS BETWEEN SCORES ON SUCCESSIVE DAYS REGARDLESS OF LENGTH

Variables	Trials		Time	
	r	P.E.	*r*	P.E.
1st day *vs.* 2d	.88	.017	.90	.014
3d day *vs.* 4th	.86	.020	.86	.020
5th day *vs.* 6th	.88	.017	.92	.012

These data fail to show any effect upon reliability from practice.

In a further attempt to determine the effect of practice upon reliability, the trials required to learn the odd- and even-numbered syllables were correlated for each of the six days. Correlations of .96, .94, .97, .96, .97, and .96 were obtained for days 1, 2, 3, 4, 5, and 6 respectively. These coefficients fail to show any advantage from practice, so far as reliability of scores is concerned. The reader's attention is called to the fact that these correlations contain in a large, but constant, spurious factor. Furthermore the writers do not regard this method of correlating odd- and even-numbered syllables as yielding true measures of reliability. The case does not seem to be comparable to the practice of correlating odd and even test items.

(3) *The Effect of the Degree of Learning upon the Reliability of Nonsense-Syllable Scores*

The relation between the degree of learning and the reliability of the scores was studied by correlational procedure. This was done by correlating the trials required to learn comparable lists to the following degrees: (1) all right twice in succession, (2) all right, (3) all right but one, (4) all right but two. Table V shows the results.

TABLE V

THE RELATION BETWEEN DEGREE OF LEARNING AND RELIABILITY OF SCORES

Degree of Learning	Trials		
	r 6 *vs.* 6	*r* 12 *vs.* 12	*r* 18 *vs.* 18
Right but two	.42 ± .064	.74 ± .035	.81 ± .026
Right but one	.57 ± .051	.74 ± .035	.83 ± .024
All right	.61 ± .048	.73 ± .036	.77 ± .031
All right twice in succession	.64 ± .046	.75 ± .034	.80 ± .027

These results indicate that, in the case of 6-syllable lists, the greater the degree of learning required the more reliable are the scores. The reliability coefficients of the 12- and 18-syllable lists seem to be unaffected by the degree of learning. Coefficients were not computed for scores of lists

learned to the point of three consecutive errorless repetitions. There were only two subjects who failed to achieve the third correct repetition after having achieved the second. Consequently the correlations would be practically the same as those under the condition 'all right twice in succession,' Table V.

These results confirm those given above relative to the relation of length of list to reliability. All these correlations, regardless of degree of learning, indicate greater reliability for the longer lists. The correlations also show considerable stability.

(4) *Further Consideration of Reliability of Nonsense-Syllable Scores*

It seems that there can be no better measure of reliability than constancy of results. There are two questions involved. The first has to do with variability of the individual from one experiment to another. The second deals with variability from one group to another. For the purpose of study the subjects were divided by chance selection into smaller groups of 38, 19, and 6 or 7 as shown in Table VI. These results are in

TABLE VI

VARIATION IN MEANS AND P.E.'S OF GROUPS OF SUBJECTS SELECTED AT RANDOM

S's	(1) 18	(2) 18	(1) 12	(2) 12	(1) 6	(2) 6
1–38....	15.2 ± .37	13.9 ± .42	10.6 ± .41	9.9 ± .36	5.8 ± .21	5.6 ± .19
39–76....	14.5 ± .49	13.5 ± .53	10.4 ± .39	10.3 ± .37	6.1 ± .21	5.6 ± .19
1–19....	15.7 ± .54	15.3 ± .65	11.1 ± .67	10.3 ± .49	5.7 ± .27	5.8 ± .29
20–38....	14.8 ± .48	13.1 ± .45	10.1 ± .46	9.6 ± .53	5.8 ± .32	5.3 ± .24
39–57....	13.8 ± .59	12.7 ± .66	10.1 ± .48	9.5 ± .41	5.5 ± .21	5.1 ± .21
58–76....	15.2 ± .21	14.2 ± .86	10.8 ± .62	11.0 ± .58	6.8 ± .34	6.1 ± .45
1– 6....	14.7 ± .90	16.2 ± 1.13	10.3 ± 1.04	9.5 ± .32	5.7 ± .44	5.7 ± .30
7–12....	15.2 ± .75	14.2 ± 1.25	10.7 ± 1.26	9.8 ± .87	5.5 ± .41	6.5 ± .34
13–18....	17.0 ± .93	15.6 ± .91	12.1 ± 1.10	11.3 ± .58	6.0 ± .54	5.6 ± .66
19–24....	13.3 ± .79	12.7 ± .51	9.7 ± .93	9.3 ± 1.03	4.8 ± .37	4.7 ± .44
25–30....	16.7 ± .67	11.8 ± .70	9.0 ± .67	8.7 ± .67	6.3 ± .72	5.5 ± .38
31–36....	15.5 ± .89	14.4 ± .85	11.3 ± .72	10.7 ± .86	6.3 ± .44	5.7 ± .40
37–43....	15.0 ± .87	15.0 ± 1.31	11.0 ± .75	9.7 ± .64	5.5 ± .31	4.8 ± .41
44–50....	13.2 ± .74	12.5 ± .80	10.2 ± .69	10.7 ± .70	5.8 ± .30	6.3 ± .21
51–56....	13.0 ± .87	10.9 ± .79	9.1 ± .92	8.4 ± .65	5.1 ± .60	4.1 ± .16
57–63....	16.5 ± 1.93	15.2 ± 1.85	11.5 ± 1.23	11.0 ± 1.32	6.5 ± .75	6.0 ± 1.27
64–70....	13.0 ± .35	12.5 ± .75	9.2 ± .55	9.3 ± .54	6.5 ± .21	6.0 ± .22
71–76....	16.0 ± 1.37	14.8 ± 1.71	11.6 ± 1.12	12.4 ± 1.20	7.3 ± .61	6.3 ± .42
Grand Average	14.9	13.8	10.5	10.1	5.9	5.6

terms of trial scores. The object was two-fold: (1) to see with what constancy the second performance of a group could be predicted from its first performance, and (2) to determine the variation between groups selected by chance.

Inspection of Table VI shows satisfactory correspondence between performances upon comparable lists. The average difference between the scores on the first and second 18-syllable lists is 1.1 trials, that between the two 12-syllable lists is .4 of a trial, and that between the 6-syllable lists is .3 of a trial. These differences may be said to be due to practice, since differences in difficulty of lists of the same length were controlled. Other chance factors ought to cancel each other so that the constant improvement shown on the second list of each pair may be regarded as due to practice. If one were to attempt to predict the second score of a pair from the first there would be a constant error equal to the practice effect. In the case of the 18-syllable lists, for example, in order to compute the error due to practice, one should subtract 1.1 (practice effect) from the first score and determine the difference between the product obtained and the second score. When the first score = 15.5 and the second = 14.4, 15.5 − 1.1 − 14.4 = 0, the error of prediction. The following observations represent errors of prediction of a group of subjects from the previous performance of the same group. The larger the group, if selected by chance, the smaller the error ought to be. With a sufficiently large group, if there be no constant errors, there ought to be no error in prediction. When the subjects were divided into two groups of 38 each the average error in predicting the mean score on the second list from the mean score on the first list when correction is made for practice is .15, .30 and .10 of a trial for the 18-, 12-, and 6-syllable lists respectively. When the 76 subjects are divided into four groups of 19 each the average error of group prediction is .36, .32 and .30 of a trial for the respective lists. Dividing the subjects into 12 groups of 6 and 7, the average error in group prediction, corrected for practice effect, is .89 (median = .32), .46 and .46 trials for the 18-, 12- and 6-syllable lists. In these computations the sign is ignored. These data indicate fair

predictability of performance of groups as small as 6 and 7 subjects when their previous performance is known. They likewise suggest that nonsense-syllable scores are fairly reliable.

Finally attention is called to the similarity between scores of different groups on the same lists. Inspection of Table VI shows that most of the group averages correspond fairly closely to the grand average of each list. None of the group averages for the 38-subject groups varies from the grand average by as much as three times their P.E. In the 19-subject groups, two of the 24 means deviate from the grand averages by as much as three times their P.E. There are 72 small groups of 6 and 7 subjects. There are 8 cases in which the group means deviate from their respective grand averages by three or more times their P.E.'s. On the basis of a single random selection of these small groups these deviations are not to be regarded as extravagant.

Inspection of Table VI will show a wide spread between the scores of the subjects who participated in this experiment. Similar correlations between scores of a less heterogeneous group should be lower, other conditions remaining constant.

Summary

The correlations between trial scores of comparable lists show greater reliability for the longer lists. Coefficients of .61, .73, and .77 were obtained for the 6-, 12-, and 18-syllable lists respectively.

Correlations of .68, .87, and .62 were obtained between the paired time scores of the 6-, 12- and 18-syllable lists. It is not known whether the comparatively low coefficient of .62 for the 18-syllable lists indicates that the reliability of time scores breaks down with the longer lists or whether it is due to chance factors. Other measures of reliability indicate that the time scores of the 18-syllable lists are more reliable than those of the 12-syllable lists.

Correlations between the trials required to learn odd- and even-numbered items show significantly greater reliability for the 12- and 18-syllable lists than for the 6-syllable lists.

They show no advantage of the 18-syllable lists over the 12-syllable lists.

The coefficients of variation and the P.E.'s indicate significantly greater reliability for the longer lists in all cases except when 12-syllable trial scores are compared with 6-syllable trial scores.

Correlation coefficients were computed between the scores of the first and second days, between the third and fourth days, and between the fifth and sixth days. There was no tendency for the coefficients to vary consistently either up or down. Coefficients were computed between the trials required to learn the odd- and even-numbered items for each of the six days. The results failed to show that reliability is increased by practice.

The foregoing results indicate that, in the case of the 6-syllable lists, the greater the degree of learning the more reliable are the scores. The reliability of the longer lists is unaffected by the degrees of learning considered in this experiment.

Generally speaking, nonsense-syllable scores may be said to be fairly reliable. This is indicated by the obtained correlation coefficients of .60 to .80 and by the rather close correspondence between the first and second scores of even small groups of subjects, when practice effect is discounted.

(Manuscript received February 9, 1933)

A NEW TEST OF LATERALITY

BY C. VAN RIPER

University of Iowa

Introduction and Procedure

Laterality of function has received a great deal of attention in recent years and seems due to be productive of much further research. The extent of the interest in problems concerned with handedness, eyedness, and sidedness is well attested by Downey's recent survey of the literature.[1] In the investigations summarized by this reviewer many different tests of handedness were employed. Since the present study is concerned with the presentation of a laterality test of different character it would seem well to enumerate briefly the tests which have been more commonly used. A representative selection of these may be listed as follows: steadiness of aiming, speed of tapping, strength of grip, throwing, writing, paper cutting, peg-board manipulation, balancing, and mirror tracing. It should be pointed out that in all these tests the differentiation of right from left sidedness involves accuracy, speed, and strength. These, however, are precisely the factors most susceptible to training and environmental influences. If, as certain studies [2, 3, 4] seem to indicate, there is a distinction between central laterality and peripheral sidedness, the latter being more subject to environmental interference, then speed, strength, and accuracy are the factors which should be minimized in laterality tests.

It would seem that an adequate laterality test should fulfill at least the following requirements: (1) it should be

[1] Downey, J.: Laterality of function. *Psychol. Bull.*, 1933, **30**, 109–142.

[2] Peterson, G. M.: A preliminary report on right and left handedness in the rat. *J. Comp. Psychol.*, 1931, **12**, 243–250.

[3] Neuman, H. H.: Studies of human twins, II. Asymmetry reversal or mirror imaging in twins. *Biol. Bull.*, 1928, **55**, 298–315.

[4] Jasper, H. H.: A laboratory study of diagnostic indices of bilateral organization in stutterers and normal speakers. *Psychol. Monog.*, 1932, **43**, 72–168.

able to show striking differences and little overlapping between extreme right and left handed groups; (2) it should have high reliability; (3) it should be capable of showing degrees of laterality if such exist; (4) it should avoid as far as possible the skills which are susceptible to previous environmental training.

The present test employs the simultaneous drawing with both hands of kinesthetic, visual, and script patterns on opposite sides of a vertical board. Preliminary investigation indicated that in such a situation the preferred or dominant hand drew the normal pattern while the non-dominant hand drew its mirrored counterpart. This seemed to hold true for both right and left handed individuals. Ambidextrous individuals sometimes produced right handed mirroring, sometimes left handed mirroring, and often no mirroring at all, drawing in the last mentioned instance the normal pattern with both hands.

In the attempt to evaluate the above mentioned observations a testing program was initiated which may be described as follows: 2500 handedness questionnaires which had been administered to freshman classes at the University of Minnesota and the University of Iowa, and to 250 pupils of the upper grades and high school of the Champion, Michigan, Public Schools were examined in the effort to select from this sampling three groups of subjects: thoroughly right handed, thoroughly left handed, and thoroughly ambidextrous individuals. Subsequently these groups were administered another handedness questionnaire, a modification of that devised by Koch,[5] the reliability and validity of which had been checked in her investigation.

These questionnaires were scored according to the formula:

$$\text{Laterality index} = \frac{R + E/2}{N},$$

where R and E equal the number of items answered 'Right'

[5] Koch, H.: A study of the nature, measurement, and determination of hand preference. *Genet Psychol. Monog.*, 1933, 13, 117–218. The questions used in our questionnaire were the twenty items listed on page 177 and items 10, in Part I, items 8 and 10 in Part II, items 3, 4, 10, 17 in Part III, and items 3, 11, from Part IV—A on pages 128–131.

and 'Either' respectively, and *N* equals the total number of items encircled. An individual who encircled all the L's would therefore have an index of 0.00, while one who encircled all the R's would have an index of 1.00.

Three groups were finally selected. The right handed group was composed of 40 subjects, the average of whose laterality indices equalled .98, individual indices ranging from .97 to 1.00. The average laterality index for the 40 members of the left handed group was .10, the range being from .00 to .18. Two criteria were used in selecting the 30 ambidextrous subjects: (1) the subject must have an index of from .25 to .75; (2) the subject must have encircled at least 20 percent of the 'Either' items. The latter requirement insured an ambidexterity in preference for individual items. The average laterality index for this group was .51.

In view of the recent interest in the relationship between mixed laterality and stuttering, a fourth group of 30 stutterers was added. The median of their indices was .90.

The apparatus and procedure used in the investigation were simple. A bimanual stylus (Fig. 2) was designed so that the blindfolded subject could learn the kinesthetic pattern with both hands working as a unit. The subject grasped the T-shaped handle with both hands but extended the forefingers into the two cups near the point of the stylus. Holding the stylus in this manner he traced the kinesthetic pattern (Fig. 1) which consisted of a reversed Y-shaped groove cut in a board until he declared he knew the pattern well enough to draw it. At least five tracings were required of all subjects. Still blindfolded, and using the same stylus, he then drew the pattern on the table-top. If successful, he was required after one retracing of the groove to draw it again but this time reduced to approximately one third size. This was necessary to insure a clear knowledge of the pattern as such. If unsuccessful he was given more tracing trials until he did know the pattern.

The stylus was then taken away from the subject and a pencil placed in each hand. He was then instructed as follows: "I am now going to ask you to draw the pattern,

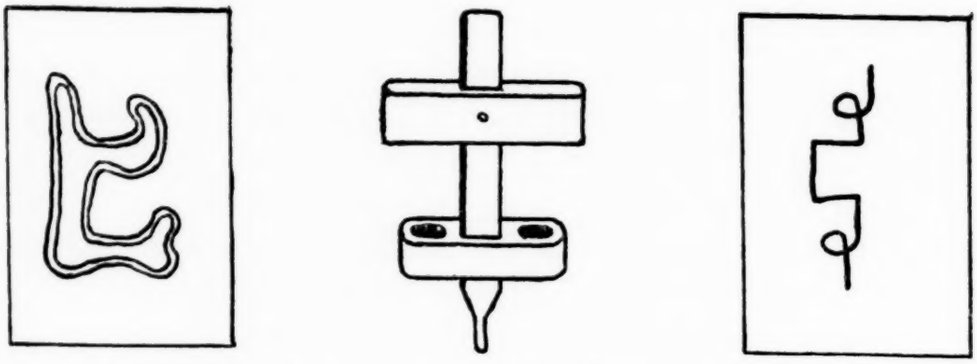


FIG. 1. Kinesthetic pattern consisting of groove cut into board.
FIG. 2. Bimanual stylus for tracing kinesthetic pattern.
FIG. 3. Visual pattern.

which you have just learned, on opposite sides of a board. You will draw it with both hands at once just as soon as I say 'Go.' Draw it as fast as you can." The points of the pencils which the subject held were then placed on opposite sides of the center of the board, and simultaneously with their contact the command to go was given. After the folded sheet on which he wrote had been removed, the subject's blindfold was taken off, and his attention called to the visual pattern (Fig. 3) which was placed on the wall or frame at an angle of approximately 30 degrees above the line of regard. This elevation prevented the subject's watching his hands. He was then instructed as follows: "You will now draw the visual pattern you see on the wall above you. As soon as I put your pencils on the board and say 'Go' you will draw it as quickly as you can. Watch the pattern on the wall all the time and trace it with your eyes as you draw it with your hands. After you have finished keep looking at the wall." Then the sheet on which he wrote was quickly removed and the blindfold again placed on the subject. He was next given the following instructions: "As soon as your pencils touch the board I shall pronounce a word. You will write it with both hands at once as quickly as you can." The word used for this script pattern was 'catch.' It should be mentioned here that the patterns and word used were chosen from a large number of similar words and patterns after preliminary research had indicated that they were both representative and convenient.

After this series of tests had been completed, the subjects

were given other tests not to be reported here which required about one half hour to complete. Then the kinesthetic, visual, and script patterns were again administered in the same way, thus providing a recheck of the original performance. A further recheck of the consistency of performance was provided by giving the tests to 51 of the subjects after an interval of from six to fourteen months. These were unselected by the experimenter who retested all who were available.

Results and Discussion

In discussing the results of the investigation as expressed in Table I, the first point to be brought out is that the test as given the first, second, or third time does in each case differ-

TABLE I

	Rt. Handed Group			L. Handed Group			Ambidextrous Group			Stuttering Group		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
Number of Subjects	40	40	23	40	40	17	30	30	11	30	30	4
Kin. Pattern, R.H. Mirroring	1	0	1	35	38	17	17	15	7	18	14	2
Kin. Pattern, L.H. Mirroring	39	39	22	3	2	0	9	11	3	9	12	0
Kin. Pattern, B.H. Mirroring	0	0	0	2	0	0	2	1	1	3	4	1
Kin. Pattern, No Mirroring	0	1	0	0	0	0	2	3	0	0	0	1
Vis. Pattern, R.H. Mirroring	1	3	2	39	39	17	14	17	6	11	10	1
Vis. Pattern, L.H. Mirroring	38	36	21	1	1	0	12	9	2	9	7	1
Vis. Pattern, B.H. Mirroring	1	1	0	0	0	0	2	0	1	8	8	1
Vis. Pattern, No Mirroring	0	0	0	0	0	0	2	4	2	2	5	1
Scr. Pattern, R.H. Mirroring	1	2	0	34	33	13	15	12	4	8	8	1
Scr. Pattern, L.H. Mirroring	39	38	23	4	5	1	11	14	4	11	13	1
Scr. Pattern, B.H. Mirroring	0	0	0	1	2	1	0	1	0	9	8	1
Scr. Pattern, No Mirroring	0	0	0	1	0	2	4	3	3	2	1	1

Legend: Kin. = Kinesthetic; Vis. = Visual; Scr. = Script;
T_1 = First test; T_2 = First retest; T_3 = Second retest.

entiate clearly between the right and left handed groups. The right handed subjects produced mirrored patterns with their left hands on all three patterns; the left handed subjects mirrored with their right hand. The amount of overlapping was small, and extremely so when compared to the overlapping found in any single test reported by previous investigators.

Ojemann,[6] for example, found no one test which differentiated between his groups; a battery of tests was needed to do so.

Of the three patterns the kinesthetic and visual patterns were slightly more discriminating than the script. This was perhaps due to the fact that the latter is more subject to environmental influence. However, several of the left handed subjects included in this investigation habitually wrote with the right hands, yet produced right handed mirror-script in this situation.

The ambidextrous group showed no consistent or typical trend either toward right handed or left handed mirroring. The former is slightly predominant indicating perhaps that many of the subjects in this group were left handed individuals who had been influenced by a right handed environment. The number of ambidextrous individuals who showed no mirroring with either hand was greater than in the other groups. The test therefore may be said to indicate not only clear cut differences between extreme groups but also that the group selected on the basis of a lack of a strong unilaterality showed a similar lack of a consistent laterality on these tests.

The speech defective group gave a performance much more akin to that of the ambidextrous group than to that of either the pure right or left handed groups. A large number of these stutterers who performed like the left handed in that they produced right handed mirror patterns were extremely right handed on the questionnaire. Of major interest however was their frequent production of mirror patterns with both hands. It is difficult to explain such a performance in any terms save those implying a confused central laterality. These subjects seldom realized that they had performed in such an extraordinary fashion, believing instead that they had drawn the pattern they had watched or learned.

Although reliability studies have seldom been run on handedness tests, it seems extremely important that they should because in this way the individual's consistency can be shown. As previously mentioned, two retests were given.

[6] Ojemann, R. H.: Studies in handedness: I. A. technique for testing unimanual handedness. *J. Educ. Psychol.*, 1930, **21**, 597-611.

All subjects took the first of these while only 51 could be obtained for the second retest approximately one year later. The stuttering group was excluded in computing the reliabilities of the tests.

The reliability data was cast in the form of contingency tables since any subject could perform in any one of four different ways. He might mirror with the right hand, or with the left hand, or with both or neither. The tables (Table II) indicate the amount of correlation between performance in the test as originally given (Test 1) and the performance in first retest or in the second retest. For example, in the table expressing the correlation between original test and first retest on the kinesthetic pattern there were 3 subjects who mirrored with both hands (B.H.M.) on the test but who mirrored with the right hand alone on the retest, 47 subjects who showed right handed mirroring on both test and retest, etc. The diagonal from lower left to upper right therefore expresses the line of perfect correlation, and it should be remembered that coefficients of mean square contingency are not identical with the Pearson product moment coefficient which expresses perfect correlation as 1.00. Perfect correlation by the contingency method is .866 for a fourfold table such as these. Therefore, the contingency coefficients for these tests, though ranging from .732 to .820 indicate a high degree of reliability not only between the first and second tests on the same day but between the first and third tests approximately one year apart.

It should be added that the reliabilities would be even higher were the ambidextrous group eliminated in computing the coefficients, since this group was responsible for more than one half of the inconsistency shown. This is to be expected because of the lack of unilateral preference so characteristic of ambidextrous individuals.

A final fact of interest is that no individual performed differently on all three of the patterns in retesting and only five individuals (of whom three were ambidextrous) performed differently on two of the three. This indicated perhaps that a battery of several patterns of each type should be used in any clinical application of the tests.

TABLE II

RELIABILITY OF TESTS AS INDICATED BY CONTINGENCY TABLES AND COEFFICIENTS

	Kinesthetic Pattern First retest					Kinesthetic Pattern Second retest			
	R.H.M.	L.H.M.	B.H.M.	No M.		R.H.M.	L.H.M.	B.H.M.	No M.
No M..	0	0	0	2	No M..	0	0	1	1
B.H.M..	3	1	0	0	B.H.M..	0	0	0	0
L.H.M..	3	48	0	1	L.H.M..	0	25	0	0
R.H.M..	47	3	1	1	R.H.M..	24	0	0	0

Contingency Coefficient .743 Contingency Coefficient .816

	Visual Pattern First retest					Visual Pattern Second retest			
	R.H.M.	L.H.M.	B.H.M.	No M.		R.H.M.	L.H.M.	B.H.M.	No M.
No M..	0	0	0	2	No M..	0	0	0	1
B.H.M..	1	0	2	0	B.H.M..	0	1	0	0
L.H.M..	6	45	0	0	L.H.M..	0	22	1	0
R.H.M..	52	1	0	1	R.H.M..	25	0	0	1

Contingency Coefficient .820 Contingency Coefficient .772

	Script Pattern First retest					Script Pattern Second retest			
	R.H.M.	L.H.M.	B.H.M.	No M.		R.H.M.	L.H.M.	B.H.M.	No M.
No M..	0	2	1	2	No M..	0	0	0	0
B.H.M..	0	0	1	0	B.H.M..	1	0	1	0
L.H.M..	3	54	0	1	L.H.M..	2	27	0	4
R.H.M..	43	2	1	0	R.H.M..	14	1	0	1

Contingency Coefficient .768 Contingency Coefficient .732

	All Patterns First retest					All Patterns Second retest			
	R.H.M.	L.H.M.	B.H.M.	No M.		R.H.M.	L.H.M.	B.H.M.	No M.
No M..	0	2	1	6	No M..	0	0	1	2
B.H.M..	4	1	3	0	B.H.M..	1	1	1	0
L.H.M..	12	147	0	2	L.H.M..	2	74	1	4
R.H.M..	142	6	2	2	R.H.M..	63	1	0	2

Contingency Coefficient .756 Contingency Coefficient .741

Legend: R.H.M. = Right Hand Mirroring; L.H.M. = Left Hand Mirroring; B.H.M. = Both Hands Mirroring; No M. = No Mirroring.
A Contingency Coefficient of .866 = perfect correlation

The tests as described have the advantage of simplicity. They are not scored in terms of speed, strength, or accuracy. They are reliable and show little overlapping between extreme groups. But the objection may well be taken that they are not able to show differences in degree of laterality. However, a modification of the test by means of an apparatus capable of shifting the planes of simultaneous writing from the horizontal to the vertical, and of measuring the critical angle at which mirroring first appears, has been employed in a second investigation and will soon be reported.

Summary

A laterality test employing the simultaneous drawing of various types of patterns on opposite sides of a vertical board was shown to differentiate groups of extremely right-handed, left-handed and ambidextrous individuals. A group of stutterers, predominantly right-handed in peripheral activities, performed similarly to the ambidextrous group. The test was shown to have high reliability for the right- and left-handed groups.

(Manuscript received January 27, 1934)